

'Improving Legal Reasoning using Bayesian Probability Methods'

Queen Mary University of London

Daniel Robert Howard James Berger



Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Date of submission: 5th December 2014

Statement of Originality

I, **Daniel Robert Howard James Berger**, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date:

Details of collaboration and publications:

1. N. Fenton, D. Berger et al, **“When ‘neutral’ evidence still has probative value: implications from the Barry George Case”**, Science & Justice, vol. 54, 2014, pp. 274-287.
2. N. Fenton, D. Berger et al, **“Response to ‘On the use of the likelihood ratio for forensic evaluation: response to Fenton et al’”**, Science & Justice, vol. 54, 2014, pp. 319-320.

Abstract

A thesis which explores the possibility of introducing Bayesian probability methods into the criminal justice system, and in doing so, exposing and eradicating some common fallacies. This exposure aims to reduce miscarriages of justice by illustrating that some evidence routinely relied upon by the prosecution, may not have as high a probative value towards its ultimate hypothesis of 'guilt' as has been traditionally thought and accepted.

For Dasha and Mum and Dad

Contents

Statement of Originality.....	2
Abstract.....	3
Contents.....	5
1 CHAPTER I: Introduction.....	9
About the author:	17
2 CHAPTER II: Conditional Probability.....	19
2.1 Introduction:	19
2.2 Frequentist v Subjective probability:	20
2.3 Axioms and theorems of probability:	21
2.3.1 PROBABILITY POINT 1:	21
2.3.2 PROBABILITY POINT 2:	22
2.3.3 PROBABILITY POINT 3:	22
2.3.4 PROBABILITY POINT 4:	22
2.3.5 PROBABILITY AXIOM 1:	22
2.3.6 PROBABILITY AXIOM 2:	23
2.3.7 PROBABILITY AXIOM 3:	23
2.3.8 PROBABILITY THEOREM 1:	24
2.3.9 PROBABILITY THEOREM 2:	24
2.4 Conditional probability:	25
AXIOM 4:	26
2.5 Conditional probability in criminal trials:.....	27
2.6 Bayes' theorem:	28
2.7 Probative value:	29
2.8 The 'likelihood ratio' (LR):	29
2.9 Event trees:	31
2.10 Bayesian networks (BNs):	34
2.11 Conclusion:.....	37
3 CHAPTER III: DNA and the Prosecutor's Fallacy	38
3.1 Introduction:	38

3.2	The Prosecutor's and Defence fallacies:	38
3.3	The random match probability:	41
3.4	Probabilities are not frequencies:.....	44
3.5	'Average' (not 'exact') numbers of matching profiles:	46
3.6	Binomial distributions:.....	48
3.7	Conclusion and recommendations:	52
4	CHAPTER IV: Digest of Legal Cases	54
4.1	Introduction:	54
4.2	People v Collins (USA), 1968:	55
4.3	R v Deen, 1994 and R v Gordon (Michael), 1995:.....	57
4.4	R v Adams (Denis), 1996:	58
4.5	R v Doheny & Adams (Gary), 1997:.....	59
4.6	R v Adams (Denis) No.2, 1998:.....	62
4.7	Regina v Sally Clark, 2000:	63
4.8	R v Dallagher, 2002:	65
4.9	R v Clark (Sally) No.2, 2003:	66
4.10	Michael Pringle v The Queen, 2003:	66
4.11	Regina v Kelly Gray, 2005:.....	67
4.12	R v George, 2007:.....	68
4.13	People v Puckett (USA), 2008:	70
4.14	R v T, 2010:.....	71
4.15	C v The Queen, 2011:.....	74
4.16	Conclusion:.....	75
5	CHAPTER V: Bayes and the Law	77
5.1	Introduction:	77
5.2	Background:	77
5.3	The early objections to the use of Bayes' theorem in court:.....	81
5.3.1	(i) That an accurate and/or non-overpowering prior cannot be devised:.....	88
5.3.2	(ii) That not all evidence can be considered or valued in probabilistic terms:.....	92
5.3.2.1	The Wigmorean method:.....	100
5.3.2.2	The key difference between the Wigmorean method and Bayes' theorem	101

5.3.3	(iii) That some evidence cannot be mathematically valued and therefore be inserted into the model, due to the existence of ‘soft variables’:	109
5.3.4	(iv) That no probability value can ever be reconciled with ‘Beyond all reasonable doubt (BARD)’:	117
5.3.5	(v) That due to the complexity of cases and non-sequential nature of evidence presentation, any application of Bayes’ theorem would be too cumbersome for a jury to use effectively and efficiently:	123
5.3.6	(vi) That probabilistic reasoning is not compatible with the law, for policy reasons:	128
5.4	Conclusion:	130
6	CHAPTER VI: Priors	132
6.1	Introduction:	132
6.2	Assessment of the prior:	133
6.3	What is a ‘prior’?	135
6.4	The value ‘K’:	137
6.5	How is a prior formulated?	139
6.6	Can the Likelihood Ratio (LR) be in any way ‘divorced’ from the prior?	142
6.7	DNA testing:	146
6.7.1	Dror and Hampikian’s DNA analysis test:	149
6.7.2	Low template DNA (LTDNA)	150
6.8	Combining evidence with different RMPs:	152
6.9	Combining RMP evidence with ‘propensity’ evidence:	153
6.10	Combining RMP evidence with location evidence:	161
6.10.1	Homogeneity:	167
6.10.2	Is homogeneity even relevant?	168
6.10.3	Priors and likelihood ratios for reference class evidence do not seem to exist in some cases:	170
6.11	DNA and the Crown Prosecution Service:	178
6.12	Conclusion:	178
6.13	Recommendations:	179
7	CHAPTER VII: Problems and limitations with the ‘Lone Likelihood Ratio’ approach	183
7.1	Introduction:	183
7.2	The likelihood ratio (LR):	183
7.3	R v George:	188

Absence of evidence:	193
7.4 R v Sally Clark:	195
7.5 Conclusion and recommendations:	198
8 CHAPTER VIII: The Full Code Test.....	200
8.1 Introduction:	200
8.2 The Full Code Test for Crown Prosecutors:.....	200
8.3 The threshold:.....	201
8.4 Submission of 'no case to answer':.....	205
8.5 Remedies available to the wronged defendant:.....	207
8.5.1 An abuse of executive power:.....	208
8.5.2 Failing to obtain evidence:.....	208
8.6 Conclusion:.....	210
9 CHAPTER IX: Overall Conclusions.....	213
Acknowledgments:	217
Bibliography	218

1 CHAPTER I: Introduction

In probability theory, Bayes' theorem (BT) can be derived from more basic axioms of probability, specifically conditional probability. With the Bayesian interpretation of probability, the theorem expresses how a subjective degree of belief in a given hypothesis should rationally change to account for evidence. Bayes' theorem is named after Thomas Bayes (1701–1761), who first suggested using the theorem. His work was posthumously read at the Royal Society before being further developed by Laplace, who first published the modern formulation in *'Théorie Analytique des Probabilités'* (1812).

In a criminal trial, a 'finder of fact' (a jury or a judge/magistrate sitting alone – although, in this thesis the term 'jury' will supplant 'finder of fact' for the sake of brevity) observes evidence to support either the prosecution or defence cases (or 'hypotheses') and returns a verdict. It is therefore logical that BT, which is designed to update beliefs on a given hypothesis after observing evidence, should be a good fit with criminal trials.

'Trace evidence' links a defendant with the scene of the crime; whether by eyewitness evidence of the defendant's presence there, a latent fingerprint, or blood stains, or footwear marks, or DNA. In recent years the use of DNA matching has exponentially increased and is now a crucial evidential tool in many cases brought by the Crown Prosecution Service (CPS). Where there is a crime scene DNA trace, and the identity of the suspect is not known, the method of presentation of this type of evidence is explicitly probabilistic and incredibly complex, with much room for error and misinterpretation of the weight of evidence, by not only laypersons, but also legal practitioners, and even in some cases, scientists, who are purportedly experienced in dealing with probabilistic evaluation. Due to these commonplace errors, and with the criminal justice system's constant striving for transparency in the trial process, it would be logical to presume that the use of explicit probabilistic reasoning in court would become more popular.

Indeed, attempts HAVE been made over the years to reconcile the criminal trial process with explicit probabilistic reasoning methods, including the introduction of BT, but these efforts have met with little success. In fact, BT is now less accepted by the legal community than ever, following the Court of Appeal decisions in cases such as *R v Doheny & Adams* [1997]

[1] and *R v T* [2010] [2], which has seen an almost blanket ban on any use of BT in the criminal justice system, except in extremely narrow circumstances, and nearly always only when DNA match evidence is involved.

In fact, the elevation of DNA, by the legal community - and some key members of the mathematical community - to a type of evidence which has almost uniquely probabilistic qualities, has meant that probabilistic reasoning in the courtroom has become somewhat of a niche discipline, instead of the usual method of combining all evidence in complex cases - which is the firm stance taken by the many supporters of BT. This thesis explores the reasons why this is so, and seeks to answer the many criticisms levelled at the use of BT in court. At the same time, the thesis will demonstrate that correctly applied BT is the only logical way to combine evidence and present it to juries, thereby becoming a crucial aid in the decision-making process. The main problem with the application of BT is the fallacy committed when attempting to combine evidence across 'types'. These types, (of which trace evidence is one, as is geographical reference class evidence – called 'location evidence' in this thesis – and propensity evidence, among others) cannot be easily combined without losing context, denoted in this thesis by the hidden value '*K*', across the model. The fallacy that these types of evidence can be combined under the currently adopted method, is identified and exposed in Chapter 6.

Further, there is no formalised approach to the decision-making process used by the CPS in deciding whether or not to bring a case to trial - despite the CPS guidelines stating that a legal test exists. The problem is that without a formalised approach to the existing test, there are no proper means of redress to the defendant. Therefore, a Bayesian approach to the current test provides a logical and rational framework to assist in helping make the trial process more transparent and accountable, from the pre-trial evidence evaluation stage to the in-trial evidence presentation stage.

For these reasons, the research hypothesis is:

'In criminal cases with multiple pieces of evidence, a Bayesian approach to evidence evaluation and presentation must be used, to increase accuracy and transparency in the pre and mid-trial decision-making processes'.

Chapter 2 explains the fundamental differences between the ‘frequentist’ and ‘subjectivist’ approaches to probabilistic reasoning. The reason that this is discussed so early in the thesis is that at the very heart of the legal debate as to why BT should be banned by the legal system for most types of evidence, is the notion that some types of evidence are objectively certain - and therefore amenable to expressly quantified probabilistic reasoning processes, such as BT - while others are not. Thereafter, the probability axioms are discussed in detail, in order to prove the validity and logical progression of thought process which leads naturally to the conditional probability axiom and the introduction of BT.

Despite this thesis being aimed at non-mathematicians, any thesis which deals with the subject matter of probability must explain the foundation principles of how probabilistic reasoning works and is expressed, in the traditional mathematical way. There may be a number of mathematic notations to encounter, but these are fairly simple and straightforward.

After explaining the inner workings of BT, there is a short discussion on the general complexity of probabilistic reasoning, with illustrations provided to give a visual representation of the enormity of the problem in trying to calculate the many variables inherent in even the simplest cases with the least amount of evidence. The solution to this complexity will be provided by the use of Bayesian Networks (BNs), for which a definition and examples are provided.

Chapter 3 begins to explore one of the most recorded of all probabilistic fallacies (and one which almost certainly increases the risk of ascribing too much weight to the prosecution case) - thereby increasing the risk of an unfair guilty verdict if the judge and jury fail to identify the error: the ‘Prosecutor’s Fallacy’. Worryingly, the Court of Appeal has begun to allow committals of the fallacy in some situations, and this thesis explores the reasons why this happens, and the dangers of allowing the practice to continue.

There is a discussion, using an in-depth analysis of Bayesian reasoning, on how the courts may have been misled into believing that some of these ‘safe’ situations may arise, and who may have assisted the courts in reaching their conclusion. The discussion in this thesis, on this point, has never been put to the legal community, and it is this author’s view that recent cases permitting a ‘threshold’ of an allowable prosecutor’s fallacy is dangerous and potentially damaging to the safety of the guilty verdicts of the defendants in each case, where

the appellate courts' reasoning has remained unchallenged. This chapter also introduces the innovative 1/World Population (1/WP) prior which will be discussed in great depth later in Chapter 6.

There are many types of probabilistic fallacies routinely committed in the course of presentation of evidence – and others, such as the 'base rate neglect', the 'jury observation fallacy' and the 'missing variable fallacy', have been well documented [3] – but the focus of this thesis is the recommendation of BT for use in criminal trials, and therefore a discussion solely of the 'Prosecutor's Fallacy' will suffice to make the necessary points to support the general use of BT for all cases.

In Chapter 4, the legal cases referred to in this thesis are listed chronologically with a summary of their facts and key probabilistic errors highlighted. This list is not supposed to be a comprehensive study of every probabilistic fallacy and error presented to a jury in court, but merely a collection of cases which best serves and encapsulates the research hypothesis.

One of the main problems with discovering faulty probabilistic reasoning in court is that while express probabilistic reasoning is easy to identify, there is also little doubt that implicit probabilistic reasoning routinely takes place during trials, simply because the adversarial trial process concerns itself with persuading a jury of the believability of either the prosecution or defence version of events – ie which is the most 'probable'? However, unless the errors are identified by the judge or defence counsel after the trial (presumably if identified during trial, the error is immediately eradicated as part of cross-examination or judicial summing-up) and the error is considered as 'material' enough by the appellate court to grant an appeal, the errors will never see light of day. This is because magistrates' court (where the majority of cases are heard) and Crown court proceedings are not recorded and made public as a matter of course, although the appeal judgments are. Therefore, unless the case is appealed, any probabilistic errors become overlooked and forgotten.

Chapter 5 aims to pinpoint the precise reason why the Court of Appeal in cases such as *R v Doheny & Adams* [1997] and *R v T* [2010] has arrived at its almost blanket refusal to allow explicit quantified probabilistic reasoning in criminal trials. In order to do this, a potted

history of the development and early criticisms of the use of courtroom probabilistic reasoning has been provided.

As with Chapter 4, the Chapter 5 collection of legal cases, scholastic viewpoints and academic theory is not meant to be comprehensive, but is designed to show the main objections to BT and the way that these objections have been answered by BT's supporters. One may suggest that the original, most wide-ranging and vehement attack on courtroom mathematics came from Laurence Tribe in his 1971 paper '*Trial by mathematics: Precision and ritual in the legal process*' [4], which was written as a criticism of the prosecutor's presentation of his case in *People v Collins* [5] - a case where multiple probabilistic errors were committed and led to an unquestionably unsafe guilty verdict for the defendants.

Since Tribe's paper, the debate concerning academic support for and against his rigid opposition to BT, has raged constantly throughout the past forty-three years, largely between those who believe that probabilistic reasoning (albeit implicitly) is commonplace in legal trials and must be formalised, and those who believe it has no place in the courtroom at all. At the time of writing this thesis, the BT naysayers have certainly triumphed thus far – resulting in the Court of Appeal's firm ruling against its routine use in the UK.

Although the substance of the debate has not moved incredibly far for the best part of a half-century, the science of identification evidence certainly has, with the use of DNA match evidence becoming a commonly-used, modern, prosecution tool. The use of DNA matching in many cases necessarily entails the use of explicit probabilistic reasoning, in the form of random match probabilities (RMPs), to present the evidence to juries, and this, with renewed vigour, has ignited the debate about the best way to present and combine DNA evidence with other types of evidence in the case.

In recognition of the complexity and high-risk approach to mathematically-untrained laypersons presenting probabilistic arguments to other mathematically-untrained laypersons, the Nuffield Foundation in London has supported a consortium of academics who have between them compiled a series of four practitioner user guides to presenting explicit probabilistic arguments in court. The Foundation's recommendations to use the non-quantitative Wigmorean method of evidence presentation for most types of evidence, apart from DNA matching, instead of the quantitative, and arguably more fit-for-purpose BT,

echoes many of the sentiments present in the Court of Appeal's stance against BT. Chapter 5 exposes and explains these similarities in a novel way, and seeks to dispel some of the more irrational objections, using Bayesian inference arguments and reasoning.

The general structure of Chapter 5 summarises and follows Tribe's five main objections to the use of courtroom mathematics and answers each of these objections using modern academic viewpoints in support. The recent work of Norman Fenton and Martin Neil [3], and others, in exploiting software for Bayesian networks as a means of reducing the complexity in Bayesian calculations, is a strong argument to counter one of Tribe's main original 1971 objections. While the science of DNA match testing has advanced exponentially in recent years, so too has the computing capabilities of software programmes like Agenarisk and Hugin which have been created and developed to meet the task.

Chapter 6 continues the earlier discussion of the ' $1/WP$ ' prior, previously mentioned in Chapter 3, and deeply discusses the reasoning behind the prior for DNA evidence matching. There has been a propensity for some influential members of the mathematical community to forget that probability theory in legal trials must reflect the beliefs of the jury in order to protect the jury's autonomous fact-finding role. Any other method of introducing BT to the criminal justice system carries the risk of subjecting defendants to a 'trial by mathematics' – a result that no-one advocates.

However, in delving deeply into the mechanics of the random match probability (RMP), and in particular explaining the problems with a mathematician assigning priors to a jury member, there exposes a fundamental issue with the RMP – that of its actual meaning. Currently, the RMP is routinely used by expert witnesses, and endorsed by the courts, to assign a particular evidential weight (or 'probative value') to DNA match evidence – usually in favour of the prosecution hypothesis. The idea is that the smaller the RMP, the unlikelier the evidence would be observed as a mere coincidence (the 'random match'), and therefore the more probative it is towards the prosecution case.

As the science of DNA matching has advanced and become more reliable, cases are being settled and verdicts of guilt are becoming increasingly reliant on DNA 'random match' evidence. However, what may have not been adequately explained to the jury in many cases is how what a 'random match' actually entails. In effect, the jury needs to know exactly how

many people in the world population, apart from the defendant, are likely to match the DNA trace found at the crime scene. By exploring the number of potential matches, we have our RMP, and therefore our probative value.

However, up until now, the expert witnesses, led by practitioner guides, such as those compiled by the Nuffield Foundation, and by Buckleton et al [6], have been presenting the RMP without much instruction to the jury as to how many people, apart from the defendant, may share the same evidential traits – no doubt a key prosecution/defence proposition. In fact, by the CPS's own guidance [7] there is no advice provided on how to present the evidence in a way which the jury can relate to - thereby leaving the defendant open to the risk of the prosecution blinding the jury with DNA evidence, the asserted importance of which cannot be effectively diminished by any other evidence in the case.

The CPS, recognising this danger, has recommended that no case should be brought on the basis of a DNA RMP alone, and recommends that DNA should be somehow combined with other evidence in the case. Chapter 6 explains how the current method of combining RMPs with other types of evidence (alibis, records of previous criminal activity, eyewitness evidence etc) in the case is almost impossible without a formalised approach provided through BNs. Further, Chapter 6 also explains how the unaided jury left to surmise on how to combine the evidence, are almost certainly going to give undue weight to the DNA evidence and therefore increase the risks of a miscarriage of justice.

In Chapter 7, the likelihood ratio ('LR') is doubted as a means of presenting evidence without the use of a full Bayesian application, such as would be provided by BNs. Until now, many influential members of the mathematical community have recommended to juries the use of the LR alone to show the probative value of evidence in cases such as *R v George* [2007] [8] and *R v Clark* [2000] [9]. However, as Fenton, Berger et al [10] have discussed in their paper '*When 'neutral' evidence still has probative value: implications from the Barry George Case*', the LR is an unreliable means of illustrating the true probative value of single pieces of evidence, without placing them within a full Bayesian model to show how the evidence fits with the prosecution case as a whole.

One of the main reasons why the advocates of the 'lone LR' have not recommended the full Bayesian approach to evidence presentation, is the reluctance to assign priors to a jury

member, as this would admit that conditional probabilistic reasoning is a subjective science, and therefore would lead to ‘uncertain’ results. However, Chapter 6 should dispel this objection from its foundation, and Chapter 7 further explains why the LR, if used without the full Bayesian model, is an unreliable method of calculating single pieces of evidence.

In the case of George, the LR was used to ‘prove’ that a single piece of firearm discharge residue (FDR) had neutral probative effect, thereby persuading the judge at the defendant’s retrial to exclude the evidence. In the case of Clark, the LR was used to show the difference between the probabilities of a mother murdering her two children against the two children dying of natural causes. In both George and Clark the LR was used without the full Bayesian interpretation of the prosecution case. If it had, the model would have shown that regardless of whether a single piece of evidence is neutral or has probative effect towards either a specific prosecution or defence hypothesis (ie that the defendant was the source of a piece of evidence or not, or whether there are many cases in the general population of natural infant deaths), this does not mean that the evidence is not probative towards the ‘ultimate’ prosecution hypothesis of ‘guilt’ or innocence’. Only a full formalised Bayesian approach introducing all of the evidence in the case together with their many variables and nuances can do this.

One of the main reasons for the difficulties in the case of Clark was that a single piece of evidence of other crimes was used to convict, without consideration of any alternative hypotheses. This has brought up issues of mutual exclusivity/exhaustivity (MEE) when considering the hypothesis pairings within a LR – something which has been highly contentious between those who believe that these pairings should be mutually exclusive (absolute and true negations of each other), without necessarily being exhaustive (exhausting all possibilities within the sample space). Chapter 7 explains how ensuring that MEE is preserved is the only means of maintaining the two opposing ‘stories’ within a Bayesian model and therefore showing the true probative value of evidence in relation to the ultimate hypotheses. After all, if the question of the defendant’s guilt is not the ultimate consideration of the jury, then what is?

In Chapter 8, BNs are recommended for use by the CPS before a case is brought to trial to evaluate, in probabilistic terms, the chance of success. Chapter 8 explains that the current test,

designed to weed-out unmeritorious actions, is not fit for purpose, and that there is no fair and effective means of redress for wrongly-tried defendants.

Currently the threshold for bringing an action is based on an explicitly probabilistic ‘more likely than not chance of conviction’ test, but that this is based on the individual prosecutor’s subjective decision after considering the prosecution and defence evidence in the case. At the moment, this decision is not open to effective scrutiny, due to its subjective nature, and therefore the only way for a wrongly-tried defendant to seek redress is to successfully defend him/herself in court – a state of affairs fraught with risks of miscarriages of justice due to a jury’s possible natural biases against defendants who are on trial.

Chapter 8 explains that a formalised Bayesian approach to pre-trial evidence evaluation is the best way to ensure that the CPS’s decision to proceed, is transparent, rational and in line with the Rule of Law and the fundamental human right to a fair trial, and non-punishment without a breach of the law.

In dealing with any level of uncertainty in evidence, it is inevitable that some assumptions have to be, and have been, made. This is the first comprehensive forensic analysis of all of those assumptions, exposing some common misconceptions and inconsistencies, and in doing so, explaining where those misconceptions and inconsistencies may lead to miscarriages of justice.

About the author:

With a Bachelor of Laws (LLB Hons) degree and a Bar Professional Training Course (BPTC) postgraduate diploma, the author has a background in law rather than statistics, and, as such - having now spent a great amount of time researching the Bayesian approach to evidence modelling - is in the rarefied position of bridging the gap between the legal and mathematical worlds. His findings, which recommend the widespread use of Bayes’ theorem in court, as long as it is correctly applied, place him in the unique position of having posited a research hypothesis, and reaching conclusions in support of that hypothesis, which are contrary to other legal scholars working in statistics and the law.

2 CHAPTER II: Conditional Probability

2.1 Introduction:

This chapter focuses on the foundations of conditional probability in order to explain how conditional probability exists and how its principles can be formulated into mathematical ‘models’ which assist in identifying and reducing uncertainty in our decision-making processes.

The beginning of this chapter discusses the key philosophical differences between frequentist and subjective probability. This is of pressing and vital importance, as discrepancies in views concerning the very cornerstone of probability has arguably led to the UK courts’ blanket ban on explicit probabilistic reasoning for most types of evidence. Therefore it is vital to begin the substantive part of this thesis with a discussion of how the opposing statistical/mathematical viewpoints, discussed later in Chapter 5, have caused confusion and misinterpretation of key probabilistic principles among those members of the legal community not trained in, or familiar with, routinely dealing with them.

This chapter then continues with a formulaic, step-by-step guide to probability, necessary to prove Bayes’ theorem – the focus of the research hypothesis – and illustrating the inner workings of causal pathways which make up the ‘event tree’ that lies at the heart of the Bayesian network. This is necessary to establish, identify and illustrate the many variables which influence our decisions.

Finally, Bayesian networks are introduced as a means of reducing the complexity in large event trees when dealing with many variables. In complex criminal cases, the number of variables exponentially increases with each piece of evidence introduced to the model, making the event tree unwieldy.

This chapter also deals, for the first time, with the notion that frequentist and subjective probability are really not opposing camps, but can only exist in conjunction with each other, which is why some of the UK court decisions listed in Chapter 4, and discussed in depth in Chapter 5, are irrational.

2.2 Frequentist v Subjective probability:

Probability is used to measure our uncertainty about the existence of past or future events. We may be uncertain either about whether it rained in London on this day in 1915 or whether it WILL rain in London on this day in 2015, but, probabilistically, the approach to uncertainty for both of these events is the same. Unless we have complete knowledge of either event, both of them carry an element of uncertainty. It seems as if there would be more certainty in past events than future ones, since the future has not happened and the past has, but this is untrue. The past is recorded and the future has yet to be recorded, but the uncertainty in each state relates to how much trust we place in our knowledge of the evidence and the world around us.

When we use our knowledge to calculate the probability of past events, we use evidence (E) to build a picture of how the world was at that time. E is collated and logged and then used to reduce uncertainty in our beliefs, which we consider in a hypothesis (H). H is merely the answer to a question we may ask ourselves before E is presented, such as: 'It DID rain in London on 19th May 1915'.

When we use knowledge to calculate the probability of future events, such as H: 'It WILL rain in London on 19th May 2015', we use data of past events (E) to build a picture of how the world may be on a specified future date. Therefore, it is possible that the same database of information might be used to collate E in the same way for both past and future events on similar subject matter. Collecting data is at the heart of frequentist probability, which is defined as the probability of seeing the event again, over an infinite number of repeated experiments.

There are two main problems with databases: (1) That past performance of an event cannot guarantee future performance; and (2) That the results from the database are open to uncertainty themselves, in terms of errors or misinterpretation. For this reason there can never be complete knowledge of any event, which is why frequentist probability is founded on an element of subjective probability. Subjective probability uses sets of assumptions and personal beliefs of the world around us to make the 'leap of faith' necessary to satisfy ourselves that what we see or are told is true. Conversely, 'objective' probability pertains to

the object, independent of the observer. The legal community are uncomfortable with the notion of subjective probability, due to its imprecise nature. However, the distinction between objective and subjective probability is irrational because even a so-called ‘complete’ database of frequentist information works on a set of subjective assumptions and personal beliefs that the database is complete.

Therefore, since subjective probability needs some frequentist information to work from and frequentist probability requires a subjective ‘leap of faith’ at its very heart, the two philosophical counterpoints are really a fabrication. In fact, subjective and objective probability must both exist in the same universe in order for us to be able to make the link between the external world and our internal beliefs.

2.3 Axioms and theorems of probability:

Taking the above two statements as examples, we have shown that they are amenable to both the frequentist and subjective definitions of probability:

- (1) ‘It DID rain in London on 19th May 1915’
- (2) ‘It WILL rain in London on 19th May 2915’

Even though these statements are of similar subject matter, one concerns a past and one a future event. However, for both events, our uncertainty can be expressed in the following point:

2.3.1 PROBABILITY POINT 1:

Uncertainty about past or future events can be expressed in percentage terms.

Percentages are sometimes expressed in an ‘odds’ form. Odds are the ratio of the event happening divided by the event not happening. Therefore odds of 2/1 means that the chance of the event happening is twice as likely as it not happening. This is the same as saying that there is a 1 in 3 chance or 66.67% chance of the event happening. Odds of 1/1 is a 50%

chance (or ‘even odds’) of the event happening. This reasoning brings us to the following point:

2.3.2 PROBABILITY POINT 2:

The percentage chance of an event happening can never exceed 100%.

This point signifies that in either frequentist or subjective probabilistic reasoning, the most certain we could ever be is 100%, which means that we can never be more sure than ‘absolutely certain’ of anything. Conversely, we can never be more than 100% ‘absolutely uncertain’ of anything, which brings us to the following point:

2.3.3 PROBABILITY POINT 3:

The percentage chance of an event not happening can never be less than 0%.

This point is consistent with both frequentist and subjective probabilistic reasoning, because we can never be less than 0% certain about any event. These last two points provide us with parameters of certainty between 0 and 100% for any event, which allows us to consider the next point:

2.3.4 PROBABILITY POINT 4:

Uncertainty expressed as a percentage, if divided by 100, can be represented by a number between 0 and 1.

This last point can now be stated by the following axiom:

2.3.5 PROBABILITY AXIOM 1:

The probability of any event happening is between 0 and 1.

This axiom satisfies both the frequentist and subjective probabilistic approaches. If we think of any event as a result of experiments testing the event's parameters, we can see that the parameters make up a 'sample space' of possible results. Taking a standard deck of 52 face-down playing cards, the probability of the top card turned over being an Ace of Clubs is $1/52$, as we can assume that the deck is standard, the number of cards is 52 and each card is different. The sample space = 52 cards. Since within any sample space all of the possible outcomes must be considered, it follows that the following axiom makes sense:

2.3.6 PROBABILITY AXIOM 2:

The probability of the combined (exhaustive) events in any sample space is 1.

Axiom 2 is a natural progression from axiom 1 due to any sample space being 100% of all of the possible outcomes. It is logical to argue that in a standard deck of 52 playing cards that there could not be a '53rd' possible outcome. This axiom relates to a single event within the sample space, but in any sample space there may be more than one event, together with experiments to test its parameters, with many different outcomes. Taking the standard deck of 52 cards as an example again, we might want to know the probability of turning the top card over and seeing the Ace of Clubs, but we also might want to know the probability of turning the top card over and NOT seeing the Ace of Clubs.

Event 1: 'Seeing the Ace of Clubs'

Event 2: 'NOT seeing the Ace of Clubs'

The probability of Event 1 is $1/52$; whereas the probability of Event 2 is $51/52$. The sum of all of the events together must follow axiom 2, adding up to 1, which means that Events 1 and 2 are mutually exclusive of each other. In essence, within the sample space, Event 1 is a true negation of Event 2, and vice versa. This principle can be elucidated in Axiom 3:

2.3.7 PROBABILITY AXIOM 3:

With mutually exclusive events, the probability of either event happening is the sum of the probabilities of the individual events.

Axioms 2 and 3 when taken together prove that within any sample space, the events and possible outcomes within it must be mutually exclusive and exhaustive (MEE). This may seem simple to understand, but errors in understanding the importance of these points has led to much academic debate when dealing with evidence in criminal trials. This issue will be explored in depth, later in this thesis, in the discussion relating to R v George [8]. Using these axioms we can derive probability theorems which follow as natural progressions from them:

2.3.8 PROBABILITY THEOREM 1:

The probability of the negation of the event is equal to one minus the probability of the event.

Using the above two events, this simply means that the probability of seeing any OTHER card than the Ace of Clubs (Event 2) is $1 - (51/52)$. In effect, the probability of Event 2 is $1 -$ (the probability of Event 1); while the probability of Event 1 is $1 -$ (the probability of Event 2). Where two events within a sample space are not necessarily mutually exclusive, we can use a probability theorem to calculate the probability of the union of these two events.

2.3.9 PROBABILITY THEOREM 2:

For any two events, the probability of EITHER event happening (their ‘union’) is the sum of the probabilities of the two events minus the probability of BOTH events happening (the ‘intersection’).

This theorem is denoted thus: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

If we take the standard deck of cards again, we can consider the experiment of drawing a single card. This time instead of calculating the probability of seeing a particular card like the ‘Ace of Clubs’ we can calculate the probability of events which are not mutually exclusive. Let us say that we wish to know the probability of drawing either any Ace OR any Club:

Event A: ‘The first card drawn is an Ace’

Event B: 'The first card drawn is a Club'

The event we wish to calculate is the probability of seeing EITHER an Ace or a Club: $P(A \cup B)$.

$$P(A) = 4/52 \text{ (4 Aces in the deck)}$$

$$P(B) = 13/52 \text{ (13 Clubs in the deck)}$$

$$P(A \cap B) = 1/52 \text{ (1 Ace of Clubs in the deck)}$$

By probability theorem 2:

$$P(A \cup B) = 4/52 + 13/52 - 1/52 = 16/52$$

$$= 4/13$$

2.4 Conditional probability:

If we were to draw a card from our standard pack and then replace it, shuffle the pack and draw again, these two draws are considered independent events. If we wish to calculate the probability that both drawn cards were the Ace of Clubs we firstly need to know the size of the sample space.

Where the standard deck is 52 cards, the number of possible outcomes is 52^2 : 2704. Assuming that each outcome is equally likely, each outcome has a probability of $1/2704$. If we ask ourselves the number of possible outcomes which could result in two draws of the Ace of Clubs after replacement of the first draw, we know that this is:

$$1/52 \times 1/52 = 1/2704$$

Therefore, we can conclude that where A and B are independent events, the probability that BOTH A and B happen is equal to the probability of A times the probability of B. In other words:

$$P(A \cap B) = P(A) \times P(B)$$

$P(A \cap B)$ (also written as $P(A, B)$) is the 'joint event'.

However, where the events are not independent of each other, for example where we repeat the above experiment but do NOT replace the cards and reshuffle the pack between the two draws; we know that the number of possible outcomes will change. In this case it is no longer 2074, but 2652 (52×51).

Let us say that we are calculating the probability of drawing ANY two Aces without replacement, we know that this is:

$$4/52 \times 3/51 = 12/1652$$

Therefore, we can conclude that the probability that both A and B happen is equal to the probability A times the probability of B GIVEN A. This can be denoted thus:

$$P(A \cap B) = P(A) \times P(B|A)$$

As long as $P(A)$ is larger than 0, the above equation can be denoted thus:

$$P(B|A) = P(A \cap B) / P(A)$$

The expression $P(B|A)$ is stated as '*the probability of B 'given' A*'. This is conditional probability and is at the heart of Bayes' theorem. Therefore our next axiom is a fundamental rule of conditional probability and is expressed as follows:

AXIOM 4:

The event 'B given A has occurred' (where $P(A)$ is not equal to 0) is written as $B|A$, with $P(B|A) = P(A \cap B) / P(A) = P(A, B) / P(A)$.

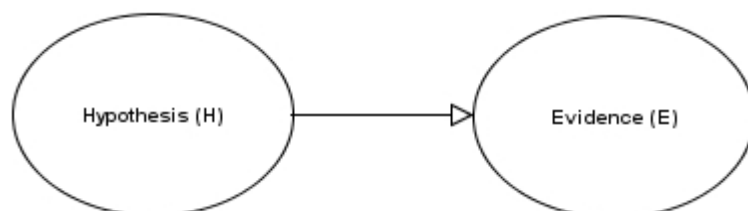
Even when frequentists assign a probability value to an event, say the probability of drawing an Ace of Clubs from a standard deck of cards, they are conditioning this on a set of assumptions about the fairness of the pack and the way the card is drawn. For this reason, ALL probability is conditional.

This thesis concerns criminal trials where there is not a simple way to establish reasonable frequentist values to all probabilities, which means that we are dealing with subjective probability at the very heart of each calculation. With probability $P(A)$ we might refer to this as a ‘degree of belief’, which will be updated with evidence of the world around us. This updating of our beliefs is something we do without thinking on a day-to-day basis, but it can be elucidated by means of ‘mathematical models’ which illustrate and visualise the many choices available before decisions are made. At the very heart of each model is a causal link between the hypothesis (H) and the evidence (E).

2.5 Conditional probability in criminal trials:

Criminal trials are a good fit for the use of conditional probability models, because the legal trial process is based upon the principle of using evidence to update beliefs about an uncertain event. Hypothesis (H) in a typical criminal trial might be ‘the defendant left a trace of DNA at the crime scene’, or it might be ‘the defendant is guilty of the crime’. Evidence (E) is used to update the jury’s beliefs about H before a verdict is expected. Therefore, the basic causal model at the heart of any criminal trial looks like this:

Fig 2.5:



If we take H (‘the defendant is guilty of the crime’) as true, then we would expect to receive evidence LINKING the defendant to the crime. Conversely, if we take H to be false we

would expect to see evidence, such as an alibi, which would DISTANCE the defendant from the crime. Using this simple causal model we can see that probabilistic reasoning involves us having a ‘prior’ belief about a hypothesis, which we denote $P(H)$, which then updates with evidence E , until we arrive at a ‘posterior’ belief $P(H|E)$, which is ‘the probability of H given E ’. Using axiom 3, We can calculate $P(H|E)$ in terms of $P(H)$, but in order to do that, we will have to know $P(H \cap E)$ - which will be rarely available. Instead, since we WILL know the likelihood of the evidence $P(E|H)$, we can compute the posterior $P(H|E)$ of the hypothesis H using Bayes’ theorem. Priors will be discussed later in much depth in Chapter 6.

2.6 Bayes’ theorem:

Bayes’ theorem (BT) gives us a simple method of calculating $P(H|E)$ in terms of $P(E|H)$ rather than the rarely known $P(H \cap E)$:

$$P(H|E) = P(E|H) \times P(H) / P(E) \quad \underline{\text{THIS IS BAYES’ THEOREM (BT)}}$$

To prove this theorem, by axiom 4, we know that:

$$P(H|E) = P(H \cap E) / P(E)$$

...which is the same as:

$$P(E|H) = P(H \cap E) / P(H)$$

...therefore, rearranging this equation gives us:

$$P(H \cap E) = P(E|H) \times P(H)$$

...which is the same as:

$$P(H|E) = P(E|H) \times P(H) / P(E) \dots \text{which is BT.}$$

One of the most noticeable benefits of BT is that it clearly distinguishes between $P(H|E)$ and $P(E|H)$. In each case the ‘conditional’ is the value after the ‘|’ symbol. By transposing them, as prosecution lawyers have routinely done in many reported cases (as will be later identified in Chapter 3), the ‘prosecutors’ fallacy’ is committed, which can have the effect of falsely altering the apparent probative value of the evidence in favour of the prosecution hypothesis - greatly increasing the risk of a miscarriage of justice. Conversely, transposing the conditional the opposite way gives false weight to the defence case – thereby causing a committal of the ‘defence fallacy’. These fallacies will be discussed later in depth in Chapter 3.

2.7 Probative value:

Where E supports H, we say it has ‘probative effect’ towards a hypothesis – ie it is evidence which ‘proves’ the hypothesis. Where E supports the negative hypothesis ($\neg H$), we say it has probative effect towards the opposite hypothesis. Evidence can also be said to have ‘neutral probative effect’ – ie that it supports neither hypothesis.

Our belief in a hypothesis is expressed as a probability. The prior probability of a hypothesis $P(H)$, is the probability of H before we observe any evidence. When there are two mutually exclusive hypotheses H and $\neg H$, the greater our belief in one, the less our belief in the other, since $P(H) = 1 - P(\neg H)$ - as was discussed in ‘probability theorem 1’ above. When we observe evidence E we revise our belief in H (and similarly $\neg H$) to our posterior $P(H|E)$. If the posterior probability is greater than the prior probability then it makes sense to say that the evidence E supports the hypothesis H, because our belief in H has increased after observing E. And if our belief in H has increased then our belief in $\neg H$ must have decreased since they are mutually exclusive explanations for the evidence, E. So, in such situations, it is both natural and correct to say that the evidence supports H over $\neg H$. The bigger the increase, the more the evidence E supports H over $\neg H$.

2.8 The ‘likelihood ratio’ (LR):

Since in a criminal trial we are only interested in whether evidence has probative effect towards either the prosecution H (‘The defendant is guilty’) or defence $\neg H$ (‘The defendant is

not guilty’) hypotheses, we are only interested in whether the evidence is true or false. Therefore, we need to consider two different likelihoods:

1. $P(E|H)$: The probability of seeing the evidence if H is true
2. $P(E|\neg H)$: The probability of seeing the evidence if H is false (and, conversely, if $\neg H$ is true)

If in a criminal case, there was an eyewitness saying he saw the defendant leaving the crime scene, we might assume that $P(E|H) = 1$, since we would expect to see this evidence if the defendant actually did leave the scene of the crime. However, we might believe that the eyewitness could have been mistaken and we might therefore attribute a probability of $\frac{1}{2}$, or 50% to $P(E|\neg H)$ on the basis that we are only half certain of the evidence.

By dividing $P(E|H)$ by $P(E|\neg H)$ we have our ‘likelihood ratio’ (LR), which represents the probative value of the evidence. $1 / \frac{1}{2} = 2$, which means that the evidence is twice as likely to be seen if the defendant is guilty than if he is not guilty. This means that the eyewitness evidence has probative value towards the prosecution hypothesis and will most certainly be put to the jury.

Of course, it is up to the jury to decide whether the evidence is enough to convict after seeing all of the evidence in the case, but the LR should be used both by the CPS in deciding whether to bring a case to trial (to be discussed in later in depth in Chapter 8) and by the jury to help in their decision-making process.

As we saw in the discussion on the ‘odds form’ of BT, the probability of an event divided by its negation is simply the ‘odds’ of the event. By rearranging BT into an odds equation, we get:

$$\frac{\text{Posterior}}{P(H|E) / P(\neg H|E)} = \frac{\text{Likelihood Ratio}}{P(E|H) / P(E|\neg H)} \times \frac{\text{Prior}}{P(H) / P(\neg H)}$$

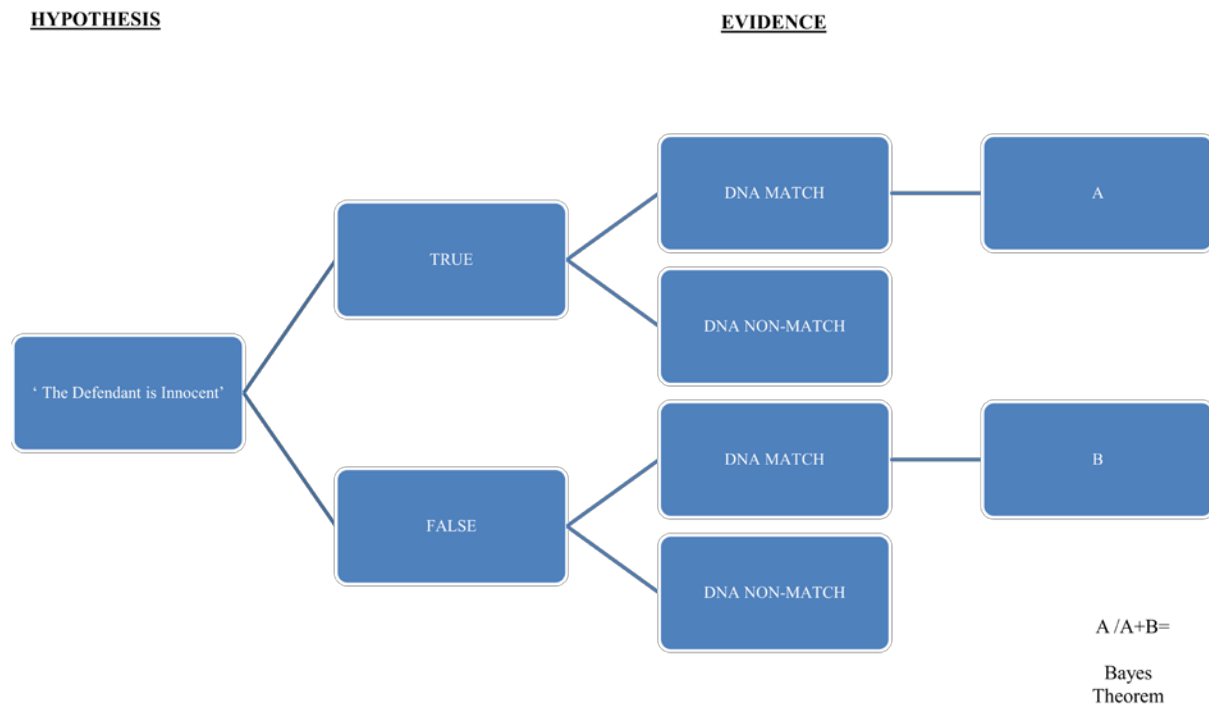
...which is the same as saying: ‘*The odds of the hypothesis given the evidence, is equal to the likelihood ratio of the evidence, multiplied by the odds of the hypothesis*’.

2.9 Event trees:

An event tree can be used to demonstrate the causal pathways of any conditional probabilistic reasoning process. By showing all of the possible eventualities in the sample space, the tree can provide calculations of all possible outcomes and will, in laborious longhand, explain the inner-workings of BT.

At the very heart of BT, when dealing with trace evidence of matching evidential traits, like DNA where the identity of the suspect is not known, is the idea that we will divide 'true positive' results by the combined 'true positive' and 'false positive' results to provide us with our posterior probability. The reason for this is that when H is true we would expect to see a positive match from the suspect (which the prosecution assert is the defendant), yet when H is false we would also expect some positive matches from other people in the world population (which is the opposite case that the defence is asserting). In the following event tree, the hypothesis H is that 'the defendant is innocent' and represents our entire sample space, which for a piece of DNA match evidence would be approximately seven billion people in the entire world population. Our prior probability is $P(H)$, or $1/7,000,000,000$. Evidence (E) is a DNA sample left at the crime scene.

Fig 2.9(a)

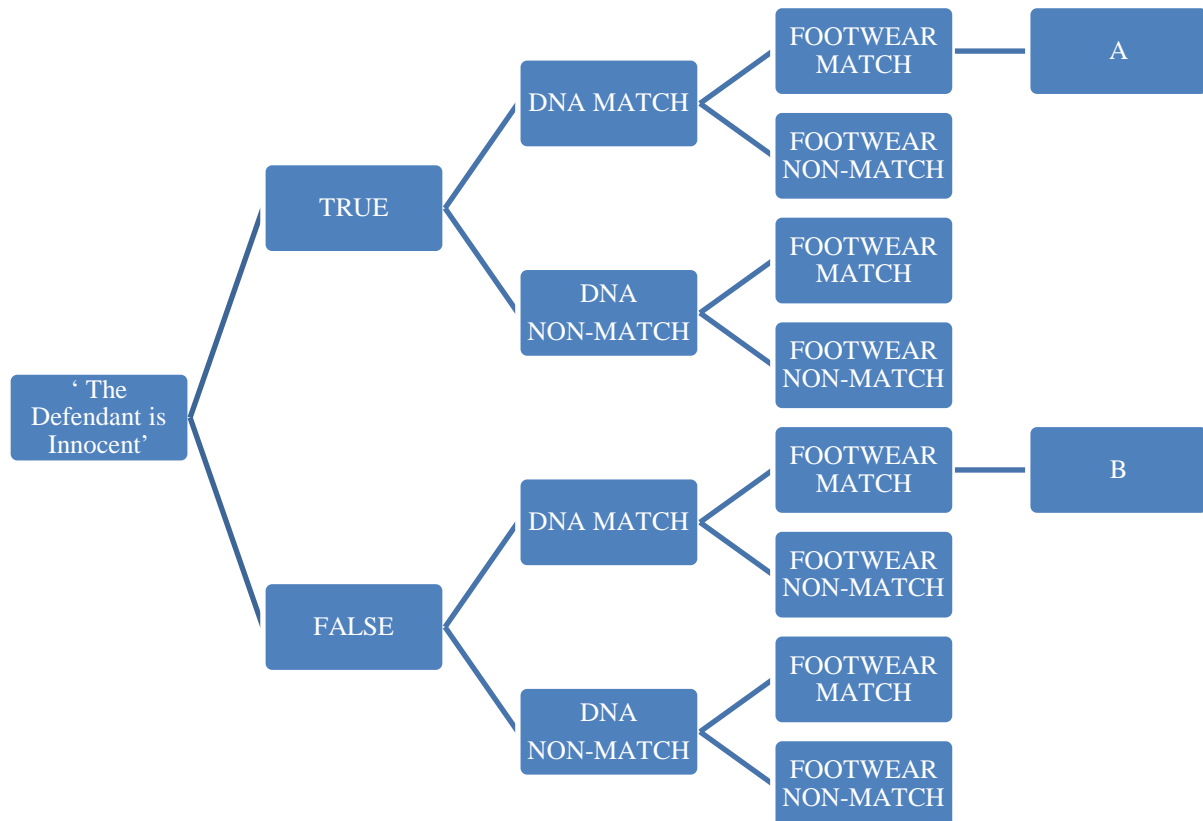


In effect, the event tree provides us with two equal and opposite states of the sample space. On one side we have true matches ending in eventual branch ‘A’ and false matches ending in eventual branch ‘B’ (we don’t bother considering the DNA non-matches as they conclude with a probability of ‘0’ and have no part to play in either prosecution or defence cases). In a case where DNA is stated to have a random match probability of 1/1,000,000,000, which means that the matching DNA profile can be found in every one billion people in the world population, there will be a ‘true’ match – ie the suspect’s, and a number of ‘false’ matches – ie where innocent people may possess the same matching evidential traits.

So, for a single piece of evidence, the event tree is fairly small with only four eventual branches at its widest point, and therefore fairly easy to negotiate and then calculate the posterior. However, criminal cases almost never consist of a single piece of prosecution evidence (in fact, for policy reasons, the Crown Prosecution Service (CPS) routinely refuses to bring cases simply on a single piece of DNA evidence and nothing else), instead relying on combining evidence to make its case. Where evidence is combined in a single tree, the number of branches exponentially rises. So for a case where there are only two pieces of

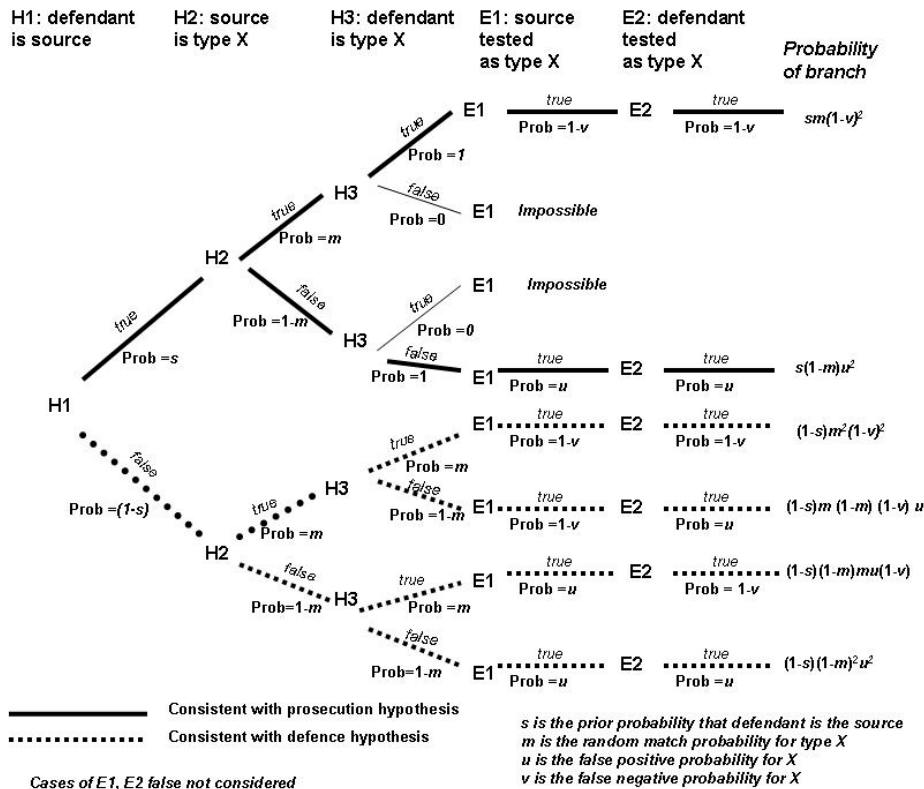
evidence, the event tree starts to become very large indeed. In the following example DNA evidence is combined with footwear mark evidence found at the scene of the crime.

Fig 2.9(b)



As can be seen here, with only two pieces of evidence introduced to the model, we now have eight branches to contend with, and many calculations to make. Fenton et al [11] do not accept that with even a single piece of DNA evidence that assumptions can be made of the certainty (authenticity) of the evidence without making the uncertainty transparent in an eight-branch event tree. The following event tree is a representation of the variables inherent within a single piece of evidence, which aims to illustrate that some fairly basic assumptions made by prosecution and defence lawyers, and then potentially passed onto juries, may need to be considered in detail.

Fig 2.9(c)



The complex nature of this supposedly simple piece of evidence is exacerbated when other evidence is then introduced to the model. An eight-branch event tree would then exponentially grow to 16, then 32 (and onwards...) individual branches, in order to provide the necessary aid to the jury's decision-making task. However, all of this unnecessary complexity can be avoided by using Bayesian networks (BNs) instead.

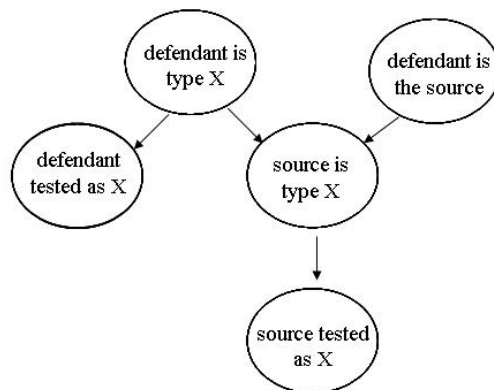
2.10 Bayesian networks (BNs):

BNs have been around for a long time. In 1991, Ward Edwards [12] explained the fundamental idea behind a BN in simple terms, understandable to non-mathematicians: 'A chance node in an influence diagram can be thought of as a four-storey building seen from its top. The top level contains the name of the node and the arcs linking it to other nodes. The second level contains a list of the mutually exclusive and collectively exhaustive states that the node can be in. The third level contains the probabilities of these states. If node C has arcs coming into it from A and B, then the probability of each state that node C can be in

must be assessed once for each possible combination of states that nodes A and B can be in. The three top levels of any chance node are accessible to, and under control of, the user. The fourth level, inaccessible to the user, contains the machinery that implements Bayesian inference'. Crucially, it is the 'fourth level' containing the 'inaccessible' computer code which simplifies the whole process and makes BNs usable by juries, lawyers and judges, who are normally untrained in explicit probabilistic reasoning.

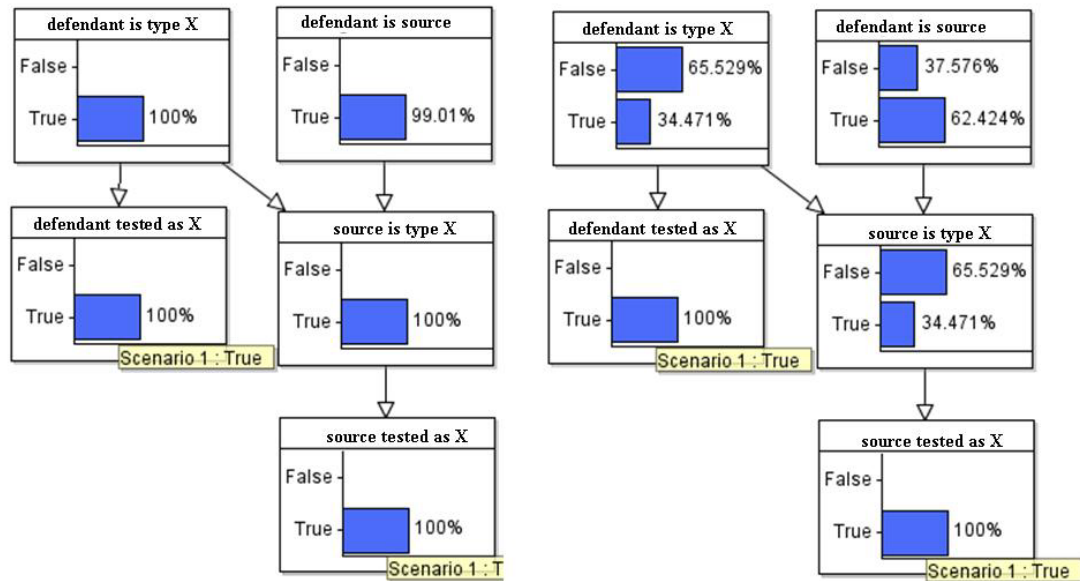
If we convert the complex eight-branch event tree at figure (?) above, the much simpler BN looks like this:

Fig 2.10(a)



...and after the probability values have been inserted into the 'third level' of the BN nodes (these two BNs in figure 2.10(b) below are shown with different sets of assumptions about the error probabilities):

Fig 2.10(b)



Fenton and Neil [3] recently define a BN as ‘...an explicit description of the direct dependencies between a set of variables. This description is in the form of a ‘directed graph’ and a set of ‘node probability tables’ (NPTs). The ‘Directed Graph (or structure of the BN) consists of a set of nodes and arcs. The nodes correspond to the variables and the arcs link directly dependant variables. An arc from A to B encodes an assumption that there is a direct causal or influential dependence of A on B; the node A is then said to be a ‘parent’ of B. We also insist that there are no cycles in the graph (so, for example, if we have an arc from A to B and from B to C, then we cannot have an arc from A to C. This avoids circular reasoning. ‘NPT’ – Each node A has an associated probability table of A, called the NPT of A. This is the probability distribution of A given the set of parents of A. For a node A without parents (also called a ‘root node’) the NPT of A is simply the probability distribution of A. An example of a NPT is shown below:

Fig 2.10(c)

H	False	True
False	0.1	0.5
True	0.9	0.3

NB. The simple illustrative NPT shown at Fig 2.10(c) sits ‘behind’ the causal $H \rightarrow E$ model at Fig 2.5, with the ‘False’ and ‘True’ values having been manually placed into each node. After the model is run, the completed BN is in the format as shown in Fig 2.10(b) above.

2.11 Conclusion:

As can be seen here, ‘Bayesian probability methods’ are really just a means of charting uncertainty in any given hypothesis. Since a BN is a graphic representation of our decision-making process, it is useful as a tool to engender debate about the weight of single pieces of evidence and how pieces of evidence combine to support either the prosecution or defence hypotheses. This debate can be used at the pre-trial evaluation stage of the criminal trial process – discussed in depth in Chapter 8 – to ensure that only meritorious cases are brought, or at the in-trial presentation stage of the process – discussed in depth in Chapters 6 and 7 – to ensure that juries understand the amalgam of evidence and dependencies between pieces of evidence in complex cases.

Therefore, the importance of using BNs is not simply to provide a formalised illustration of complex probabilistic reasoning, but it is a means to eradicate fallacies, rectify errors in communication and promotes transparency in the decision-making process. These attributes all have the potential to reduce miscarriages of justice – an aim which **MUST** be in the interest of the legal community, to say nothing of the rest of society.

The legal community should not concern itself with whether the underlying mathematical reasoning supports the introduction of the widespread use of Bayes’ theorem and BNs in court, but should concern itself more with the values placed into the model. This aspect of the problem is dealt with in detail in Chapter 6.

3 CHAPTER III: DNA and the Prosecutor's Fallacy

3.1 Introduction:

This chapter introduces an in-depth analysis of one of the key and most commonly committed fallacies in criminal cases: the prosecutor's fallacy. The chapter begins with a definition of the fallacy, together with an explanation of how the fallacy is committed. The chapter then continues with an explanation of why the fallacy can occur when trace evidence, such as unattributed DNA samples left at a crime scene - presented as a random match probability (RMP) – can and routinely do lead to a committal of the fallacy.

The chapter then continues with a forensic examination of the RMP to establish whether the UK courts' current stance, seen in key recent decisions such as *C v The Queen* [13] and *R v Kelly Gray* [14], that there can be an 'acceptable committal of the prosecutor's fallacy', is logical.

The courts' potentially highly controversial stance is an area which has never been discussed, researched or refuted by any legal or mathematical scholar to date, and which, if found to be untenable, could leave the guilty verdict decisions mentioned in this chapter - together with the many which are left unnoticed, unreported or knowingly following the precedent set out by the Court of Appeal - open to serious questions as to their safety.

The discussion in this chapter is vital in supporting the research hypothesis and provides the UK courts with a clear framework in which to understand the issues and controversies its recent decisions have exposed.

3.2 The Prosecutor's and Defence fallacies:

The fallacy of the transposed conditional, or in criminal trials, the 'prosecutor's fallacy' [15] confuses $P(H_d|E)$ with $P(E|H_d)$, thereby overvaluing the prosecution case against the defendant, as will be seen in many of the cases in Chapter 4. This transposition can also overvalue the defence case, known as the 'defence fallacy' where $P(H_p|E)$ is confused with $P(E|H_p)$. Taking the example of a piece of DNA evidence with a random match probability

(RMP) of 1/1 million, which means that the same evidential traits can be found in ‘one in every million people in the world population’, transposing the conditional in this case will have the following effect:

Hp: ‘That the defendant is the source of E

Hd: ‘That the defendant is not the source of E’

E: DNA with a RMP of 1/1 million

$P(E|Hd) = 1/1$ million, which in other words is: ‘*the probability of seeing the evidence given that the defendant is not the source is 1/1 million*’,

is confused with:

$P(Hd|E) = 1/1$ million, which in other words is: ‘*the probability that the defendant is not the source, given the evidence is 1/1 million*’.

$P(E|Hd)$ tells us that there are approximately 7,000 potential evidential matches in a world population of 7 billion people, therefore the probability of the defendant being the source is actually 1/7,000, while $P(Hd|E)$ tells us that the probability of the defendant not being the source is 1/1million – grossly overstating the prosecution case.

Buckleton et al [6] have recognised the risk of a jury misinterpreting the value of a prosecution or defence case when transposing the conditional, but have argued that in some cases the effects of this fallacy are negligible, or in some cases nil. It is argued that ‘*the transposition is of no consequence if the prior odds are in fact 1*[- which, although unclear here, means that $P(H) = P(\neg H)$]. *This is because the answer arrived at by transposition and the ‘correct’ answers are the same in this circumstance. The only issue occurs if the prior odds differ from 1. If the odds are greater than 1 then the transposition is conservative. For a high likelihood ratio (a low RMP) the practical consequences are negligible* [- this is the key error which the UK courts have unfortunately accepted]. *The practical consequences, if they occur at all, are for lower likelihood ratios and where there is little ‘other’ evidence against the defendant, or where there is evidence for the defendant*’. Only when $P(Hd) = P(\text{not } Hd)$ (i.e. a 50-50 prior which we will argue later in this thesis is never correct) is $P(E|Hd)$ equal to $P(Hd|E)$ – i.e. only in the special case is the effect of transposing $P(E|Hd)$ with $P(Hd|E)$

irrelevant. In all other cases, contrary to what Buckleton says transposition will lead to the wrong results. Let us say that the prior for DNA evidence is $1/WP$ (world population of approximately 7 billion people), which means that the hypothesis pairing for this evidence is:

Hp: 'That the defendant is the source of the DNA'

Hd: 'That someone else in the world population, alive at the time of the crime, is the source of the DNA'

As a misinterpretation of Buckleton's advice, the UK courts would argue that if the RMP for the evidence is very low (let us assume $1/WP$ or lower), then the effect of a committal of the prosecutor's fallacy is nil, because:

$P(E|Hd) = 1/7$ billion, which in other words is: *'the probability of seeing the evidence given that the defendant is not the source is 1/7 billion'*

...means the same as:

$P(Hd|E) = 1/7$ billion, which in other words is: *'the probability that the defendant is not the source, given the evidence is 1/7 billion'*.

In other words, where the RMP is $1/7$ billion and there were around seven billion people in the world population at the time of the crime, that the probability of the defendant *not* being the source of the evidence is $1/7$ billion, which means that there could only have been one suspect – the defendant. The Nuffield Foundation [16] supports Buckleton's and the UK courts' view that it is possible to legitimately transpose the conditional under certain conditions: *'The conditional is legitimately transposed through the application of Bayes' Theorem. Illegitimate transpositions arise through confusion and are always unjustifiable. Whether replicating the classical "prosecutor's fallacy" or some variation on source probability error, illegitimate transpositions adopt the flawed logic of thinking that "If I am a monkey, I have two arms and two legs" implies that "If I have two arms and two legs, I am a monkey".'*

In the conjoined appeals of *R v Doheny & Adams* [1], Lord Phillips sitting in the Court of Appeal stated: *'the more remote the random occurrence ratio, the less significant will be the*

adoption of the “Prosecutor's Fallacy”, until the point is reached where that fallacy does not significantly misrepresent the import of the DNA evidence.’ In the case of Doheny the RMP of the DNA evidence was 1/40 million, while in Adams the RMP was 1/27 million – in neither case was the blatant committal of the prosecutor’s fallacy said by the Court of Appeal to be a significant risk to the soundness of the convictions.

In more recent cases such as *R v Kelly Gray* [14] and *C v The Queen* [13], Lord Phillips’ judgement was interpreted by the Court of Appeal to provide a RMP ‘threshold’ by which the committal of the prosecutor’s fallacy was deemed to be safe. In Gray the RMP was 1/81 million, while in C the RMP was 1/1 billion. Neither conviction was overturned despite blatant transposition of the conditionals which doubtless went unnoticed by the juries.

The questions are, can the committal of the prosecutor’s fallacy ever be deemed ‘safe’, and can it ever be the case that transposing the conditional has no impact on a criminal trial? To answer these questions, an examination of how random match probabilities are formulated is essential.

3.3 The random match probability:

The RMP is a measure of the probativity of a piece of trace evidence – usually DNA, but could also be blood type, fingerprints or hair colour etc – where the identity of the suspect is unknown. Buckleton et al [6] call a RMP from a crime scene sample of trace evidence, a ‘*coincidence probability*’ and explains how one is formulated: ‘*Formulate the hypothesis, H_0 : the DNA came from a male not related to the suspect. We then calculate the probability of the evidence if this is true. We write the evidence as E , and in this context it will be something like: E : The DNA at the scene is type α . We assume that it is known that the suspect is also type α . We calculate the probability, Pr of the evidence, E , if the null hypothesis H_0 is true $Pr(E|H_0)$. Assuming that about 1 in a million unrelated males would have type α , we assign $Pr(E|H_0)$ as 1 in a million. Since this is a very small chance, we assume that this evidence suggests that H_0 is not true and hence is support for H_1 . In this context, we might define the alternative hypothesis as: H_1 : The DNA came from the suspect. Hence in this case, the evidence supports the hypothesis that the DNA came from the suspect’.* The RMP itself came from sampling a section of the world population to arrive at

single probabilities for single locus pairings – markers which show profiling traits along the DNA molecule, as the Nuffield Foundation [16] explains: *‘Ideally, a good sample is constituted by a “random sample” of the target population i.e. that group of individuals about whom information is sought. In a random sample, every member of the target population has an equal probability of being selected as part of the sample. One must ensure that the population from which the sample is taken (the sampled population) actually is the target population.’*

These loci, assumed to be independent of each other, are then multiplied using the product rule to formulate the RMP. Buckleton states that a *‘moderate database of ≈ 200 individuals’* is enough to provide robust estimates of the probabilities of single locus pairings. These probabilities are not frequencies, but a measure of a degree of belief in the hypothesis of a ‘match’. Curran [17] recognises that it is hard to understand how databases of 200 individuals can provide statistics with probabilities of 1/1 billion, and that the U.S. Federal Bureau of Investigation (FBI) are known to use such a small sample.

The answer is that a sampling error rate is built into the RMP to allow for an acceptable margin of error. Redmayne [18] acknowledges that it is vital to preserve the general principle of allowing the benefit of any doubt to lie with the defendant: *‘If match probabilities are to be calculated in a situation of uncertainty it is important that those probabilities are demonstrably conservative. But, given that several different standards of conservatism have been suggested, it is difficult to choose which one best accords with the norms of criminal procedure. Just how conservative should the statistical calculations used in criminal trials be?’* Curran recommends ‘credibility intervals’ – a means of reducing uncertainty by formulating a band of certainty from posterior probabilities and ignoring the higher and lower fields to provide, say, a 95% band - are used, while Buckleton suggests a number of ways that sampling errors can be calculated, including a ‘factor of ten’ rule – a general application whereby RMPs are multiplied by ten to give the defendant the benefit of any doubt. This system alters, say, a 1/1billion RMP to 1/100 million to increase the pool of suspects from around 7 to 70, thereby reducing the probability of the defendant’s culpability.

The Nuffield Foundation [19] supports the use of confidence intervals to ensure that the correct allowances are made for inadequate sampling methods: *‘At the allelic level, the question is whether several hundred allele counts are sufficient for calculating ethnically-*

stratified allele probabilities. With appropriate values for allele probabilities, it can be shown that no more than several hundred alleles are required to generate robust estimates of allele frequencies when genotype probabilities are calculated using sampling and co-ancestry allowances such as those illustrated below. However, the adequacy both of the sampling allowance method and of the number of allele counts should always have been formally assessed using a statistical method like the one reported by Curran and Buckleton.'

However, the Nuffield Foundation [16] do not support the notion of confidence intervals for 'criminal proceedings' as it would for 'social sciences' - possibly for the reason that any interval is arbitrarily and/or artificially created, thereby leading to inaccuracy in evaluating the probative value of the evidence it represents: *'...employing categorical levels of confidence leads to evidence "falling off a cliff" – i.e., it is excluded entirely - if it falls outside the chosen confidence interval, even by a tiny margin. Evidence which may be highly probative within the stated confidence interval is arbitrarily allotted a value of zero if a small change takes it outside that (arbitrarily chosen) confidence interval. Whatever the merits for social science in proceeding in this fashion, it is plainly unsatisfactory for evidence to be allowed to "fall off a cliff" in criminal proceedings, especially when it is recalled that assessments of statistical significance are merely a way of representing variation in data. Consequently, the fact that a particular estimate falls outside one's preferred confidence interval does not necessarily mean that this result is uninteresting or provides an inaccurate measure of real world events which are themselves subject to natural variation.'*

This distinction between criminal proceedings and other disciplines where DNA testing is used, must rest primarily on the notion that an unsafe criminal conviction carries a higher risk to society than a mistake might do in other spheres.

Perhaps, for this reason, the Crown Prosecution Service (CPS) [7] adopts the practice of capping all RMPs at a low probability band of 1/1billion, regardless of the expert's RMP valuation. Because a RMP of 1/1 billion means that the number of suspects in the pool is approximately 7 in the world population, this means that no juror could ever safely convict a defendant purely on the strength of the DNA evidence alone, because with more than one suspect still in existence, the prosecution can never assert that it has presented sufficient evidence so that the jury are able to meet the standard of proof necessary to return a guilty verdict.

Apart from the 1/1 billion RMP cap, the CPS have stated that DNA evidence is never used alone to convict a defendant because there is always other evidence in the case. Despite Buckleton's claim that a RMP of 1/7 billion leads to the Bayesian inference of there being only one suspect carrying the evidential profiling traits, the CPS state that it is policy to always cap RMPs at 1/1 billion, which further clouds the issue as to whether it is ever possible to infer a single match in the world population, after only having sampled 200 members of it. Fenton and Neil [3] suggest that in a pool of 10,000 suspects, where the evidence is a blood type match with a RMP of 1/1,000 that *'there are around ten suspects'*, while, similarly, Colmez and Schneps [20] state that matching thirteen loci from two separate profiles are *'considered sufficient to completely determine the identification of an individual'*. However, there are also various contradictory arguments which undermine this standpoint, mainly focusing on the difficulty of ascertaining how many 'matches' there are in the world population at any given time. To resolve this difficulty, the RMP must be examined in greater detail.

3.4 Probabilities are not frequencies:

When calculating a RMP, Buckleton et al [6] distinguishes between 'probabilities' – a measure of certainty in a given hypothesis - and 'true frequencies' – empirically gathered data containing *'a numerator and a denominator, e.g 3 in 25, where we have counted 3 particular outcomes out of a possible 25'* -, whereby *'a frequency of a genotype will be a probability only if we could conceive of carrying out an experiment of randomly sampling, with replacement, individuals chosen from the population of the world at a given instant... [and as such] we cannot directly compare our probability assignments to true values'*. The difference between the two is that a probability value does not reflect an exact expected frequency, but more of a rough estimate.

Since this is an estimate, Buckleton et al accept that *'this makes it important that these inherently untestable probabilities are assigned by the most robust methods'*, and suggests that the test for this is a *'fair and reasonable assignment of probability'* which is a more practical solution than calculating true frequencies, which would *'typically require the genetic typing of the whole population of the world, and the values would change constantly as individuals were born or died'*. A probability statement about how many actual matches a

RMP would provide (*‘typically small numbers’*), would not be the same as a frequency statement, which as Buckleton et al acknowledges might be *‘0, 1, 2 or more in [the world population]’*. Therefore, a RMP is merely a statement of certainty relating to the following hypothesis pairing:

Hp: ‘That the defendant is the source of the evidence’

Hd: ‘That someone else in the world population, alive at the time of the crime, is the source of the evidence’

Although a RMP is merely a measure of uncertainty in this pairing, it must adhere to the ‘fair and reasonable’ test governing its formulation, otherwise it is useless. Since P(Hd) refers to suspect matches ‘who were alive at the time of the crime’ within the world population, this number needs to be properly estimated. Buckleton et al state that while:

(i) *‘DNA allele proportions are expected to remain constant from one generation to the next’*;

...they also state that:

(ii) *‘...the allele probabilities in one generation would still differ slightly from the previous one, caused by the random transmission of alleles to the new generation.’*

The juxtaposition of these two propositions means that while DNA allele sampling results in a fair and reasonable assessment that a particular allele appears to be a constant feature in a section of an infinite population, its fluctuation would vary from the mean in the short term. This makes intuitive sense. Taking gender as an example, this would be the same as saying that the proportion of male-female babies remains constant at 50/50 from one generation to the next in the long term, but the probability of there being an exact 50/50 split in the very next generation is low. This argument is a species of the ‘Law of Large Numbers’ [21] whereby infinite distributions tend to regress to the mean.

The regression to the mean occurs where the sample space is infinite; much in the same way that trends in the world population are taken to be. Taking a coin flip example (where it assumed that the coin is fair); we can expect an even spread of heads and tails over the infinite term, but not over a finite term. In this way we avoid the ‘gambler’s fallacy’ –

whereby a single event, taken alone, thought to be dependent on past events actually is independent of them. In any finite future space, of say a 10 flips, the probability of seeing 10 heads is no different to seeing 9 heads + 1 tail, or 8 heads + 2 tails etc. While a run of 10 heads has a probability of 1/2,048, this is only the case before the first coin is flipped. After the first 9 flips, the results are no longer unknown, so their probabilities are 1. Reasoning that the past flips were all heads, and so must somehow influence the future flip, is what triggers the gambler's fallacy.

The same is true with DNA allele pairings. Just because a test sample formulates a probability of seeing a particular allele in a proportion of the infinite world population, does not mean that the proportions will be consistent over a finite period of time within this infinite sample group. It could be argued that sampling the world population at the exact time of the reported crime *would* allow a more precise assessment of the probability, but this would actually have the reverse effect, since any diversion from any previous test sample would provide *more* credence to the argument that the allele proportions are constant while the probabilities fluctuate periodically.

Therefore, once the very first test sample provides the probability of an allele, any future sampling of that allele will not provide a more accurate assessment of its proliferation - but will only serve to chart its diversion from the mean since the last sampling. Taking the coin flip example again, a test sample of flips can provide a fair and reasonable assessment that the coin is fair and will return an even spread of heads and tails over the infinite sample space (much in the same way that 200 test subjects is enough to provide '*robust*' estimates of allele probabilities in the world population), but any future test sample of the same coin can do nothing to change the proportions of heads and tails over the infinite future sample space – only the probability of heads and tails within that short time frame. In fact, only by constantly and infinitely testing the entire sample space, will accurate values ever be found – a measure that the CPS [7] says is '*not practicable*'.

3.5 'Average' (not 'exact') numbers of matching profiles:

Let us say that DNA has a RMP of 1/1 billion. If the world population is, say, 6.5 billion people at the time of the reported crime, we can expect between 6 and 7 matches. We cannot

be sure whether there are 6 or 7, because we don't know how many, if any, people with matching profiles has died since the test sample, nor can we say exactly when in the future a person with a matching profile will be born.

Therefore, we say that the number of matches is an 'average', much in the same way that for every week of a lottery with 14 million possible combinations, and 14 million tickets sold, we can expect, on average, 1 winner per week – as is consistent with Buckleton et al's views, there might actually be 0, 1, 2 or more winners. Therefore, the number of matches in the world population of any crime scene DNA sample, at any given time, can only ever be an average, with diversions from this average to be expected. Even taking a new test sample to formulate new probability values will not address this issue, since all it will do is confirm that the allele numbers have fluctuated since the last sample, while not addressing the key issue of whether the allele proportions have changed.

Buckleton et al suggest that it is possible to calculate a RMP small enough to provide an inference of '*a 1/100th of a suspect*', while Colmez and Schneps, despite stating that a 13 loci match might be a 'certain match', also argue that in the rape case of *People v Puckett* [22] that a probability of innocence can be calculated. This probability is based on comparing a RMP against the prior of the world population and arriving at a posterior of 1/70 in favour of the defence hypothesis. But what is a '*1/100th of a suspect*'; and what does a '*1/70 probability of innocence*' actually mean? Actually, a 100th of a suspect first and a 1/70 probability of innocence, in the context of the case of *Puckett*, means the same thing, namely that the posterior of the RMP is 1/100 and 1/70, respectively, in favour of the defence hypothesis.

Taking the example of the 1/100 posterior; the RMP needed to arrive at this value is 1/700 billion (as previously discussed in this chapter, Buckleton et al assert that it is possible to calculate RMPs as small as this, even with only 200 test subjects). The case is as follows:

Hp: 'That the defendant is the source of E'

Hd: 'That someone in the world population, alive at the time of the crime, is the source of E'

E: DNA with a RMP of 1/700 billion

LR: $P(E|H_p)/P(E|H_d) = 700 \text{ billion}$ (assuming a 0 error probability rate)

Prior $P(H_p)/P(H_d) = 1/WP$ (assuming a world population of around 7 billion people)

Bayes' theorem: $1/WP \times 700 \text{ billion} = 100/1$ (posterior) – or $1/100$ in favour of the defence hypothesis.

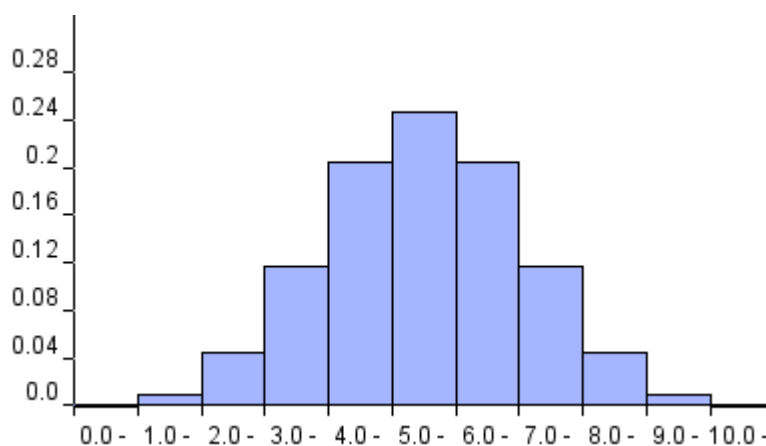
A similar calculation is used to arrive at a posterior of 70 – or $1/70$ in favour of the defence hypothesis – simply by adjusting the RMP. In a finite world population sample space, any RMP smaller than $1/WP$ provides a posterior which implies a single suspect (and in this way the identity is certain) but in an infinite world population sample space, where the number of matches is only an average, the number of suspects is implied only by means of a binomial distribution of results.

3.6 Binomial distributions:

Binomial distributions are used to model the number of successes in a small sample, *with replacement*, taken from larger populations, such as with DNA match testing.

If we wish to consider an example of flipping a fair coin, for which the probability of 'heads' is $P=1/2$, how many heads should we expect in 10 flips? Clearly, we expect to get 5 heads. But we should not always expect exactly 5 as sometimes we might get 4 or 6, and in more extreme circumstances, either 0 or 10. However, what we need to know is how far from 5 would it be reasonable to get? In other words, what is the probability that we get exactly 5 heads in 10 flips? The answer is in the following binomial distribution graph:

Fig 3.6



As we can see, at the peak of the graph, we expect 5 heads, but the distribution tails towards 0 and 10 at each end. Regardless of how many coin flips we make, the number of heads we will expect to see will always be an average 5.

If, however, we test a sample and know at the outset that we can expect an even distribution of successful results, the graph would take a very different shape, as the number of heads we would expect to see is exactly 5 from 10 flips.

Taking the lottery example again - where an average single winner per week is expected, due to the number of combinations and ticket sales corresponding exactly - the probability of a single winner for any given draw is 'on average, one'. If there is no winner this week, then next week we still expect a single winner, but *over the two weeks taken together*, we would expect to see two winners. In this way, the probability of seeing one winner per week adheres to the law of large numbers. Therefore, the posterior of 1/100 in favour of the defence hypothesis simply means that if the sample group were multiplied by a factor of 100, in the same way as making 100 separate lottery draws, we would *still* expect only one match. Of course, we might not actually see a match at all, since this is only an expectation.

Interestingly, the Nuffield Foundation [16] have come to a similar conclusion on the number of potential matches but have posited a reason based on 'familial matches': *'Even though a random match probability may be extremely small (one in ten billion, say – the world's estimated current population being (only) six billion) it does not warrant the inference that a matching DNA profile uniquely identifies an individual. Quite apart from anything else, every set of identical twins in the world has the same DNA profile – and the chances of obtaining random matches are vastly increased in relation to parents and siblings. With a random match probability of, e.g., one in ten billion and a world population of six billion, the probability that there is at least one other person with the profile is about 0.46 (and a corresponding probability of 0.54 that no-one else does). For six billion people and a random match probability of 1 in 260 billion, the probability of at least one other match in the population is about 0.02.'* However, the Foundation STILL has not expressly accepted that believing that a future match will occur from past events is simply a commission of the gambler's fallacy: *'The random match probability must not be confused with the probability of obtaining another match somewhere in the population. The random match probability is the probability of obtaining a match "in one go", not the probability that at least one other*

member of the population of interest will produce a match. The probability a particular person identified in advance will win a lottery is different from the probability the lottery will be won (by someone).’ Incredibly, this explanation is precisely the same conclusion one would reach on the gambler’s fallacy hypothesis.

The fallacy that certain matches can be made from RMPs, have been perpetrated for many years. For instance, in the 1968 U.S. robbery case of Collins [5], the prosecutor attempted to prove that a certain match from evidence could be made by combining naturally occurring traits, such as race and hair colour, which are found proliferating across the sample space. Finkelstein and Fairley [23] quite rightly refuted this, and formulated a Bayesian approach to calculating the true probative value of this type of evidence.

In fact, what Finkelstein and Fairley did *not* discuss, was the binomial distribution of ‘average’ results made from trace evidence – a point that was equally missed by Edwards [12] when he sought to provide more a more accurate Bayesian view of the evidence in Collins, than the prosecution case presented. If they had, then a problem would have been exposed – that an increasing accumulation of trait evidence can lead to smaller and smaller RMPs, which carry a heavy risk of a Bayesian misinterpretation of the weight of the prosecution case against the defendant. To illustrate this problem, let us say that the case against a defendant is as follows:

Hp: ‘That the defendant is the source of E^1 , E^2

Hd: ‘That someone else in the world population, alive at the time of the crime, is the source of E^1 , E^2

E^1 : Fingerprint profile with a RMP of 1/2,000

E^2 : Blood type profile with a RMP of 1/1,000

Using a prior of 1/world population – taken to be approximately 7 billion people - ($1/WP$), BT shows us that $P(Hp)/P(Hd) \times P(E|Hp)/P(E|Hd) = P(Hp|E)/P(Hd|E) = 1/3,500$ in favour of the defence hypothesis. Let us now introduce a new piece of evidence to the model:

E^3 : Male gender profile with a RMP of $\frac{1}{2}$

The posterior becomes $1/1,750$ (assuming all evidence types are independent of each other), signifying that are 1,750 matches. Now take a fourth piece of evidence:

E⁴: Racial profile with a RMP of $1/50$

The posterior becomes $1/35$, signifying that are 35 matches. Now take a fifth piece of evidence:

E⁵: Height profile with a RMP of $1/35$

The posterior becomes 1, signifying a certain match. The problem is that for any piece of trace evidence with an RMP, the more evidence accumulated, the lower the probability of a coincidental match. Of course, once ‘average’, rather than ‘exact’ numbers of matches are considered, the balance can shift to the opposite hypothesis to create a more accurate assessment of the probative value of the combined evidence. To illustrate this point, now take a sixth piece of evidence:

E⁶: Weight profile with a RMP of $1/10$

Our posterior $P(H_p|E)/P(E|H_d) = 10$, or $1/10$ in favour of the prosecution hypothesis. This shift to the opposite hypothesis does not mean a certain match, but shows that distribution of results has now moved into an ‘average expectation’ over a larger sample space of repeat tests. Colmez and Schneps, and Buckleton et al have shown the effect of crossing the ‘hypothesis boundary’ from prosecution case to defence case, without explaining that in neither of the hypotheses does the evidence point to a certain match. By acknowledging ‘average’ matches we can reconcile Finkelstein and Fairley’s argument that RMPs can never provide certain matches, with the idea that there is potentially infinite amounts of trace evidence that can accumulate against a defendant in an infinite world, and reject any argument that there is a threshold by which a small enough RMP leads to the Bayesian inference of a single ‘exact’ match.

By looking at where the hypothesis boundary threshold is, we can say that any combination of trace evidence supporting the defence hypothesis = a posterior of $<1/2$, while any combination of trace evidence supporting the prosecution hypothesis = a posterior of $>1/2$,

with the binomial distribution graph curving towards infinity in either direction, thereby never reaching certainty for either hypothesis. The point of parity is at $\frac{1}{2}/\frac{1}{2}$, which signifies neutrality in the case (this could also be shown as 50/50, or 1) – not a single match.

3.7 Conclusion and recommendations:

Strictly speaking, although Buckleton et al's statement that in some cases a committal of a prosecutor's fallacy will have nil *mathematical* effect, it will still have an effect on a jury's perception of the value of the prosecution case, which means that there is disconnect between some *mathematical* values and the *real world* values they represent. As has been shown in the examples here, these mathematical values can belie the fact that the crucial difference between $P(E|H_d)$ where we are dealing with 'exact' numbers of matches, and $P(E|H_d)$ where we are dealing with 'average' numbers of matches, means that the significance of a prosecutor's fallacy can easily be misconstrued due to a misunderstanding of how many suspects in the world population carry the identical evidential traits.

The problem arises where we are dealing in average numbers of matches. A jury thinking that there are exact numbers of matches could easily misinterpret a RMP of $1/WP$, or lower, to mean that the DNA evidence points to certain guilt.

The reason for this is that, much like the lottery example, we expect a single 'winning ticket' on average, but as to whether there will be exactly one winner per week, will be down to long term averages, rather than single event predictions. Therefore, any RMP must come with an express warning as to the Bayesian inferences which should be drawn from it. The Court of Appeal has obviously misunderstood the significance of 'average' matches when it ruled that there is such a thing as a 'safe' prosecutor's fallacy.

As has been argued here, there is no such thing as an acceptable threshold whereby a prosecutor's fallacy will have a negligible or nil negative impact on the defence case. The RMP thresholds in *R v Adams & Doheny* are arbitrary, and should not be seen as a target by which verdicts should or should not be considered 'safe'. In fact, it could be argued that since RMPs *are* linked to the Bayesian inferences of numbers of matches in the world population that a binomial distribution graph should be made available to juries to ensure that they fully

understand the significance of what a particular RMP actually means in terms of suspect numbers. Since experts like Colmez and Schneps have misunderstood the potential numbers of matches from RMPs, it is almost certain that laypeople on juries would make the same mistake, without this assistance.

Further, when dealing with ‘acceptable thresholds’, a further danger is that a prosecutor, knowing he has reached the Adams & Doheny RMP of 1/27 million or smaller, may commit the prosecutor’s fallacy with impunity, knowing that due to the Court of Appeal’s ruling, the convictions will be safe even if the defendant has his complaint upheld that the fallacy may have been instrumental in leading to his conviction.

4 CHAPTER IV: Digest of Legal Cases

4.1 Introduction:

The following is a list of criminal law cases which contain explicit probabilistic reasoning by prosecution witnesses and/or counsel in order to establish the guilt of the defendants. This is not an exhaustive digest of cases which contain probabilistic reasoning, but this selection carries the advantage of (a) having been reported; and (b) containing reasoning errors/fallacies/legal judgments which help illustrate and expose many of the problems encapsulated in this thesis.

This is the first ever definitive one-stop summary of the key cases which encapsulate the heart of the controversy concerning the use of probabilistic reasoning in court. For the first time, all of the salient probabilistic issues have been collated in a single document in order to identify and illustrate the necessity for a formalised approach to criminal evidence evaluation and presentation.

The cases are listed chronologically to demonstrate the history and evolution of courtroom probabilistic reasoning, and the way that the courts have reacted to its sporadic attempts at introduction, and have the key relevant issues italicised and in brackets below the names of each case. Chapter 5 will deal the reasons behind some of the rulings and the current anti-Bayes stance taken by the Court of Appeal.

What unites all of the cases in this chapter is that they feature probabilistic fallacies and errors which could have been resolved by an application of Bayes' theorem (BT). The cases of Collins, Denis Adams (Nos. 1 and 2) Doheny & Adams and T demonstrate that the courts' approach to BT has evolved from a 1968 tentative approval of courtroom mathematics, to an almost blanket ban today, following systematic errors in reasoning and presentation. The cases of Deen/Gordon, Dallager and Pringle demonstrate that, at one time, the courts recognised the importance of the prosecutor's fallacy as a potentially dangerous way of misleading a jury as to the weight of the prosecution case against the defendant, but that following the ruling by Philips LJ in the 1997 Doheny & Adams case, have now started to allow committals of the fallacy as long as it reaches a certain 'safe' threshold; and the more

recent cases of Gray and C demonstrate the application of this threshold and the faulty reasoning which led to its creation. The cases of George and Clark are included to demonstrate that the likelihood ratio (LR) should not be used without a full application of BT, to ensure that the jury are not misled as to the prosecution's 'ultimate hypotheses' of guilt/innocence. Puckett has been included to demonstrate that no trace evidence match can ever prove uniqueness of the sample – a point which was raised in the 1968 Collins case, and academic literature following the case, but never fully resolved.

4.2 People v Collins (USA), 1968:

(Prosecutor's fallacy/Inappropriate use of product rule)

Mr and Mrs Collins were arrested for robbery on the basis of eyewitness evidence, who claimed to have seen an inter-racial couple work together to steal the victim's handbag and make a getaway in their yellow Lincoln car. The policeman in charge of the case, who knew of the Collins', suspected and questioned them on the basis of their racial orientation and car ownership.

The eyewitness evidence was that the woman, who did the bag snatching, was white and wore a pony tail in her hair, and that the man was black and had a beard and moustache.

At trial:

The prosecuting attorney, Ray Sinetar, had a local mathematics professor, Daniel Martinez, give evidence to the jury that non-independent probabilities could be multiplied together to give a combined probability figure, using the product rule. Sinetar attributed the following values as '*characteristic individual probabilities*':

- A. Partly yellow automobile 1/10
- B. Man with moustache 1/4
- C. Girl with ponytail 1/10
- D. Girl with blond hair 1/3
- E. Black man with beard 1/10
- F. Interracial couple in car 1/1000

After multiplying these probabilities together and reaching a figure of 1/12,000,000, the attorney assured the jury that it could be sure beyond reasonable doubt that the Collins' were the handbag robbers. The jury convicted the Collins' and they were sentenced to prison.

On appeal:

Lawrence Tribe, a Harvard mathematician, clerked for a Supreme Court judge as Collins' appeal eventually reached the highest appellate court. Tribe advised the panel that Sinetar's submissions had no basis in sound mathematical reasoning.

The Supreme Court found that B and E could not be seen as non-independent probabilities, since the likelihood of having a beard is increased in men who wear moustaches. Also, the probability attributed to A was not made with any statistical support, and was merely the opinion of a mathematician with no experience of such things. Comment was also passed on whether C could ever be a true probability, on the basis that ponytails could be done or undone at will. F implies that the two people in the car were a 'couple', although their actual relationship is unsubstantiated. Also, even if the probability figure of 1/12,000,000 could be substantiated, this still does not mean that the Defendants were guilty beyond reasonable doubt – the number would simply represent a random match probability of coinciding events. On the above reasoning, the convictions were quashed.

Key parts of the court's judgment:

'The prosecution's introduction and use of mathematical probability statistics injected two fundamental prejudicial errors into the case: (1) The testimony itself lacked an adequate foundation both in evidence and in statistical theory; and (2) the testimony and the manner in which the prosecution used it distracted the jury from its proper and requisite function of weighing the evidence on the issue of guilt, encouraged the jurors to rely upon an engaging but logically irrelevant expert demonstration, foreclosed the possibility of an effective defence by an attorney apparently unschooled in mathematical refinements, and placed the jurors and defence counsel at a disadvantage in sifting relevant fact from inapplicable theory.'

‘While we discern no inherent incompatibility between the disciplines of law and mathematics and intend no general disapproval or disparagement of the latter as an auxiliary in the fact-finding processes of the former, we cannot uphold the technique employed in the instant case.’

‘The prosecution produced no evidence whatsoever showing, or from which it could be in any way inferred, that only one out of every ten cars which might have been at the scene of the robbery was partly yellow, that only one out of every four men who might have been there wore a moustache, that only one out of every ten girls who might have been there wore a ponytail, or that any of the other individual probability factors listed were even roughly accurate.’

4.3 R v Deen, 1994 and R v Gordon (Michael), 1995:

(Prosecutor’s fallacies)

In *R v Deen*, a rape case in which the prosecution counsel transposed the conditional $P(E|H)$ with $P(H|E)$, thereby wildly overstating the prosecution’s case, exposed and fully discussed the prosecutor’s fallacy - which has now been cited in numerous later judgments. On the basis that Deen’s appeal was allowed and a retrial ordered, an appeal of the rape case of *R v Gordon (Michael)* was heard. In *Gordon*, the forensic expert, Dr Greenhalgh stated: *‘...the probability of a [match] in two unrelated persons is less than one in 159 million... taking into account a conservative measure on defaulting, that is reduced to one in 17.1 million.’* After *Gordon*’s trial, new evidence emerged which threw doubt on the reliability of the DNA database. On appeal, it was stated that the effect of this new evidence was to throw doubt on any useful statistical evaluation of the evidence. As it could not be said with certainty what the effect of this additional evidence was, the appeal was allowed and a retrial ordered. In relation to Dr Greenhalgh’s evidence at trial, the Lord Chief Justice disparaged the use of large random match probabilities presented to juries, by stating: *‘Figures running into millions of the kind put before the jury have a dramatic quality which may well exert a strong influence upon them.’* This case was one of the first to recognise the strong influence a very small prior probability, founded on a single piece of evidence, can have on the evidence in the case as a whole.

4.4 R v Adams (Denis), 1996:

(Probabilistic reasoning and presentation errors)

The case of *R v Adams (Denis)*, concerned a rape trial in which a DNA sample was obtained from a swab taken from the victim. At trial the defence was permitted to lead evidence of BT in connection with the evaluation of the evidence in the case. The appeal, which was raised on three grounds, and which the first two grounds failed, rested solely on whether the judge misdirected the jury as to the evidence in relation to the BT and left the jury unguided as to how that theorem could be used in properly assessing the statistical and non-statistical evidence in the case.

At first instance trial:

Professor Donnelly for the defence gave evidence, by agreement, before the prosecution case was closed. He said it was logical and consistent for the jury to deal with the rest of the evidence in the case in statistical terms and for the jury to do this using the BT. He identified four types of evidence which could be evaluated in this way, namely the probability that the offence was committed by a local man (which the appellant was), the non-identification evidence, the appellant's evidence and the alibi evidence.

Professor Donnelly explained in detail the figures he used in the mathematical model and the methodology used to provide the likelihood ratios. The trial judge, in summing-up, misdirected the jury when he became confused by the expert evidence and, in particular, appeared to have forgotten the answer given by Professor Donnelly expressing the probabilities as percentages. Instead, the judge was directing the jury as to how many times it was more likely that something had occurred, yet he did not remind the jury of the formula given by Professor Donnelly in relation to the percentages.

On appeal:

The Court of Appeal held that: *‘It seems to us that the difficulties which arise in the present case stem from the fact that, at trial, the defence were permitted to lead before the jury evidence of the Bayes Theorem’* (at page 480), and later stated that: *‘...the apparently*

objective numerical figures used in the theorem may conceal the element of judgment on which it entirely depends... it seems to us that it is not appropriate for use in jury trials, or as a means to assist the jury in their task... Jurors evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them... as the present case graphically demonstrates, to introduce Bayes Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task' (at page 482). Accordingly, the appeal was allowed, the conviction quashed and a retrial was ordered.

4.5 R v Doheny & Adams (Gary), 1997:

(Prosecutor's fallacy/Inappropriate use of product rule/Failure to explain the meaning of a 'random match probability')

A year after *R v Adams (Denis)*; *R v Doheny*, an appeal (conjoined with *R v Gary Adams*) of a rape trial in which the evidence primarily relied upon by the prosecution was a blood stain left at the crime scene, was based on shortcomings in the DNA evidence and the manner in which it was presented to the jury.

Doheny, at first instance trial:

In *Doheny*, the forensic expert, Mr Dowie, arrived at a random match probability, of the blood sample, of 1 in 40million, after multiplying a 1 in 5.7million DNA loci random match probability, after a multi locus probe testing, with a 1 in 7 blood type random match probability of the Defendant's blood. In his examination-in- chief, Mr Dowie gave the answer to a leading question on the likelihood of any other person than the Defendant leaving the blood stain. His response was '1 in 40million;' thereby committing the 'Prosecutor's Fallacy.' The question should have been left to the jury to decide on all the evidence whether they were sure that it was the appellant who left the crime stain, or whether it might have been one of a handful of other persons who might exist in the United Kingdom sharing the same DNA profile.

On the conjoined appeal:

The ground of appeal, concerned whether it had been legitimate for Mr Dowie to multiply the result of the multi locus probe test with the results of the single locus probe tests, and whether there is a risk that the multi locus probe test will identify bands which are the same as those identified by the single locus probes, or in such close proximity to them that there is a likelihood that they will be linked together, so that whenever one is found the other will be likely to be present also.

On behalf of the Appellant, Dr Debenham, a molecular biologist, stated in his report: *'...as there is no way of determining in any one multi-locus probe test whether the bands identified are clustered with respect to any of the single locus probe alleles it cannot be assumed that the tests are independent'* (at page 380). The Crown accepted that the possibility of dependence existed. Under these circumstances the conviction was quashed and a re-trial would have been ordered if the Appellant had not already served half of his sentence and was eligible for parole and/or would have been prejudiced by the long period of time since his trial, in which he relied on alibi evidence, which would by now be likely inaccurate if heard again.

A submission was made by counsel for *Doheny* relating to the role of the expert in trial proceedings. The Court of Appeal stated: *'He will properly, on the basis of empirical statistical data, give the jury the random occurrence ratio - the frequency with which the matching DNA characteristics are likely to be found in the population at large. Provided that he has the necessary data, and the statistical expertise, it may be appropriate for him then to say how many people with the matching characteristics are likely to be found in the United Kingdom - or perhaps in a more limited relevant sub-group, such as, for instance, the Caucasian, sexually active males in the Manchester area... It has been suggested that it may be appropriate for the statistician to expound to the jury a statistical approach to evaluating the likelihood that the defendant left the crime stain, using a formula which gives a numerical probability weighting to other pieces of evidence which bear on that question. This approach uses what is known as the Bayes Theorem. In the case of R v Adams (Denis) this Court deprecated this exercise...We would strongly endorse that comment'* (at page 375).

Adams (Gary), at first instance trial:

Adams (Gary), concerned a forensic scientist, Mr Webber, who had given evidence that he had carried out four single locus probe tests of the match, that he had found between the crime stain and the appellant's blood sample, which had resulted in eight matching bands, producing a random occurrence ratio of 1 in 27 million.

In his examination-in-chief, Mr Webber stated: *'I can estimate the chances of this semen having come from a man other than the provider of the blood sample. I can work out the chances as being less than 1 in 27 million'* (at page 383), thereby, as in *Doheny*, committing the 'Prosecutor's Fallacy.' In summing-up to the jury, the trial judge stated in relation to a random member of the population who could possibly have left the DNA sample: *'...on the evidence of Dr Webster if you accept it and there is nothing to contradict it, not less than 1 in 27 million people. That means, I should think, I do not know what the population of the United Kingdom is but I should not think there were more than 27 million males in the United Kingdom, which means that it is unique'* (at page 384).

On the conjoined appeal:

In dismissing the appeal, the Court of Appeal stated: *'Professor Donnelly gave evidence for the appellant in this case also. The best he could do was to point out that on that ratio there was a probability of about 26 per cent that at least two men in the United Kingdom, in addition to the appellant, had the same DNA profile as the crime stain... The complainant had, quite comprehensively, identified one man - the appellant - as her assailant: the telephone call, his voice, his appearance, his clothing. When to this was added the fact that his DNA profile matched the crime stain, no jury could be in doubt that it was he who left that stain, whether the statistics suggested that there existed one other man, or 10, or even 100 in the United Kingdom with the same DNA profile'* (at page 386).

Lord Phillips, at page 379 stated, per curium: *'the more remote the random occurrence ratio, the less significant will be the adoption of the "Prosecutor's Fallacy", until the point is reached where that fallacy does not significantly misrepresent the import of the DNA evidence. Such was the position on the figures advanced by Mr Davie. Before us, however, Mr Davie's figures were not merely attacked, but significantly undermined'* (at page 379).

4.6 R v Adams (Denis) No.2, 1998:

(Probabilistic reasoning and presentation errors)

At first instance retrial:

In *R v Adams (Denis) No.2*, the previous appeal of which had resulted in a retrial; since in the first appeal there had been no argument as to whether BT was admissible and the court had only merely *per curium* suggested that it had grave doubts as to its admissibility, the defence team once again invited the jury to use BT to calculate the probabilities of the non-DNA evidence being true or false, but on that occasion provided the jury with a questionnaire to enable them to make the appropriate calculations.

On second appeal:

The three grounds of this second appeal, advanced by Ronald Thwaites QC, defence counsel, were: (i) that as the prosecution had adduced evidence using BT, then the defence should be allowed to call their own BT expert to analyse the data and reduce the numbers if found to be appropriate; (ii) that BT is logical, sound and approved by expert opinion; and (iii) that BT evidence having been admitted, the judge should have directed the jury fully and not encouraged them to apply their common sense in contradistinction to the Bayesian approach described by Professor Donnelly.

The court gave their ruling on the three grounds of appeal in reverse order. Firstly, on the third ground, the Appellant argued that the trial judge should have summed-up Professor Donnelly's evidence and did not treat the questionnaire with sufficient gravity. The court disagreed, and stated that this was unreasonable submission in light of the extensive evidence presented and the questionnaire provided to the jurors. In addition, that although BT is used as a mathematical tool to assist in the decision-making process, the judge did state that it in no way precludes the operation of what the prosecution expert, Mr Lambert' called 'a common sense approach' to the evidence. The judge stated: *'If you do not wish to use it that is your privilege and your own private decision and no one will criticise you for not using it. There is absolutely no compulsion on you to use it at all. It is there if you want to use it and follow the instructions given'* (at page 383).

Secondly, on the second ground, the court stated: *‘We would not for our part wish to take issue with that statement so long as it is applied to appropriate subject matter by persons competent to apply it. We have no reason to doubt, as is stated by a number of highly authoritative experts, that it is a sound and reliable methodological approach in some circumstances. We have, however, the gravest reservations about its use in jury trials in cases such as this’* (at page 383).

Thirdly, on the first ground of appeal, the court stated: *‘Of course, it is a matter for the jury how they set about their task, and it is no part of this Court’s function to prescribe the course which their deliberations should take... We do not consider that they will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion. We are very clearly of opinion that in cases such as this, lacking special features absent here, expert evidence should not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial’* (at page 385). For these reasons, the appeal failed.

4.7 Regina v Sally Clark, 2000:

(Prosecutor’s fallacy/Probabilistic reasoning errors/Inappropriate use of product rule)

The case of *Regina v Sally Clark* concerned the death of a mother’s two baby sons, who were born and died a year apart from each other and which, the defence proffered, occurred by Sudden Infant Death Syndrome (‘SIDS’) or ‘cot death.’ The post mortem on the first baby was carried out by Dr Williams, who noticed that there was some bruising and a damaged frenulum which would be consistent with a resuscitation attempt. A year later, Dr Williams performed a post mortem on the second child and his findings were indicative of non-accidental injury, consistent with shaking on several occasions over several days, and it was considered that shaking was the likely cause of death. In the light of this, further tests were carried out in relation to the first child and Dr Williams altered his opinion, concluding that the first child’s death had also been unnatural and that the evidence was suggestive of smothering.

At first instance trial:

At trial, it was the prosecution's case that the first baby died from smothering and the second baby from suffocation after having received a violent trauma to the spine. It was alleged that neither death could be considered SIDS, because of the existence of recent and old injuries that had been found in each case, nor was there sufficient evidence as to how they had been caused. The circumstances of both deaths shared six main similarities which the prosecution argued against the defence hypothesis that either death was natural: (1) the babies were about the same age at the time of death, namely 11 weeks and 8 weeks; (2) they were each found by the appellant unconscious in the same room; (3) both were found at about the same time, shortly after having been fed; (4) the appellant had been alone with each child when he was discovered lifeless; (5) in each case the Defendant's husband was either away or about to go away; (6) in each case, according to the prosecution, there was evidence of previous abuse and of deliberate injury recently inflicted. The defence case was that the appellant did not kill her children or do anything untoward, and that they must have died of natural causes.

The Defendant was convicted for the murder of her two children. There were thirteen days of evidence in all, and it was not until the defence expert medical evidence was called that it became clear that the medical experts called by the defence accepted that neither death was a true SIDS death. The principle of SIDS rarity became a central issue, on the basis that the prosecution anticipated that the defence might assert the rarity of two babies independently dying of the syndrome, which thereby led to Professor Meadow giving prosecution evidence – an exercise which the defence later claimed may have misled the jury into treating the possibility of SIDS as a central issue. Statistical evidence was presented at trial by Professor Meadow on the likelihood of a mother's two children dying from SIDS.

Professor Meadow's original witness statement stated that: *'Even when an infant dies suddenly and unexpectedly in early life and no cause is found at autopsy, and the reason for death is thought to be an unidentified natural cause (Sudden Infant Death Syndrome — SIDS), it is extremely rare for that to happen again within a family. For example, such a happening may occur in 1:1,000 infants therefore the chance of it happening within a family is 1:1,000,000. Neither of these two deaths can be classed as SIDS. Each of the deaths was unusual and had the characteristics of a death caused by a parent'* (at para 110). Between

1993 and 1996, the Confidential Enquiry into Stillbirths and Deaths in Infancy ('CESDI') produced a major study of 470,000 live births, which concluded that SIDS occurred in 1:8543 live births, producing a squared ratio of 1:73,000,000 for two coincidental SIDS deaths in one family. When counsel for the Crown learnt that Professor Meadow (who, by his own admission, is not a statistician) was writing a preface to the CESDI study, they used this higher figure when submitting a Notice of Additional Evidence.

On appeal:

The ground of appeal included the Appellant's submission that Professor Meadow had cited erroneous figures to support his report, that Professor Meadow's opinion as to the deaths being unnatural was wrongly founded in part on the statistical evidence cited in breach of the guidelines in *R v Doheny & Adams*, and the judge failed to warn the jury against the prosecutor's fallacy, as referred to in *R v Deen*, in relation to the use of statistical evidence.

During trial, Professor Meadow gave evidence that: *'It gives a chance of 1 in 73 million live births and in England, Wales and Scotland there are about, say, 700,000 live births a year, so it is saying by chance that happening will occur once every hundred years'* (at para 130). In summing-up the evidence, the trial judge warned the jury about the use of statistics, but did not adequately, in the Court of Appeal's opinion, clear away the 1:73,000,000 ratio as a distraction, which he should have done. However, since the statistical portion of the trial was deemed to have had a minimal impact on proceedings compared with the other adduced evidence, the appeal was dismissed and the conviction upheld.

4.8 R v Dallagher, 2002:

(Prosecutor's fallacy)

The case of *R v Dallagher* concerned an ear print left on a window at the scene of the murder of a 93 year old woman. The appeal included a ground which the defence asserted was of the prosecutor's fallacy evidence given by the forensic expert, Mr Van der Lugt, who stated when asked whether there was a comparison between the crime scene ear print and the Defendant's ear print, that he was: *'convinced that they are from the same donor'* (at para

32). A further expert, Professor Vanezis, less emphatically stated that his conclusion was that: *'it was the closest match for the overall fit of the prints'* (at para 33).

The Court of Appeal refused to grant the appeal on the ground of a committal of the prosecutor's fallacy, and Kennedy LJ commented that: *'...as any juror can appreciate, comparisons such as were made in this case cannot be expressed in terms of statistical probability'* (at para 34). However, the appeal was granted and a retrial ordered on a different ground on the basis of new available evidence. At retrial in 2003, the Prosecution offered no evidence against the Defendant and he was immediately acquitted.

4.9 R v Clark (Sally) No.2, 2003:

(Prosecutor's fallacy/ Probabilistic reasoning errors/Inappropriate use of product rule)

In Sally Clark's second appeal, the Appellant contended, on the first ground of appeal that the pathologist had withheld vital evidence supporting the defence hypothesis, and on the second ground of appeal that statistical information given to the jury about the likelihood of two sudden and unexpected deaths of infants from natural causes had misled the jury and overstated very considerably the rarity of two such events happening in the same family.

The court criticised Professor Meadow's evidence-in-chief, who stated that in order to visualise a likelihood ratio of 1:73,000,000, the jury should imagine backing the winning horse in the Grand National horse race at odds of 80:1 each year for four years running. The problem with that approach is that the rest of the evidence in the case, in light of this single assertion, is rendered insignificant, regardless of its true value. Since this second appeal was also brought on the grounds of withheld defence evidence, the court quashed the conviction on the basis that the first Court of Appeal panel would have also quashed the conviction if it had known about the withheld evidence.

4.10 Michael Pringle v The Queen, 2003:

(Prosecutor's fallacy)

At first instance trial:

In *Pringle*, the Privy Council presiding over a Jamaican rape/murder appeal, found that the expert witness, Dr Yvonne Cruickshank, and the trial judge together propounded the prosecutor's fallacy in giving and hearing evidence about the certainty of the DNA crime scene match with the Defendant's DNA when a conversation between her and the judge during trial contained the following exchange: 'His Lordship: *And you told me it was 99.999 per cent certainty?* Witness: *No. I said to you I would have to say it is with a high degree of certainty.* His Lordship: *Not 99.999...* Witness: *I said 99.999. In science we say 99.99999. It goes on. But we did not address the probability in this.* His Lordship: *Should I qualify this 99.999 now?* Witness: *Not in the context of which we spoke. It will still stand*' (at para 17).

In summing up, the judge referred to Dr Cruickshank's evidence and stated: '*So, it is based upon these results that she comes to the conclusion that the spermatozoa there came from Pringle, that it, that Pringle had sexual intercourse with the deceased*' (at para 18).

On appeal:

On the basis of the fallacy committed between the judge and the expert witness, the Privy Council quashed the conviction and remitted the case to the Jamaican Court of Appeal to consider whether or not to order a retrial.

4.11 Regina v Kelly Gray, 2005:

(Prosecutor's fallacy)

At first instance trial:

Gray concerned a DNA sample left at the scene of a drug deal, which also culminated in the Defendant facing charges of grievous bodily harm. One of the grounds of appeal related to the evidence given by Miss Doole, the forensic expert, in one of numerous fallacious exchanges with prosecuting counsel was asked: '*The sample on the jacket that you looked at and found a profile which matched the defendant's profile — one in 81 million — was a sample that was part of the blood, was it not?*' to which she answered: '*It came from that sample of staining*' (at para 20).

Also, in summing-up, the trial judge stated: ‘...*the probability of that blood coming from someone other than [the victim] or unrelated to [the victim] is of the order of one in a billion, that is one in a thousand million...*’ (at para 20).

On appeal:

In refusing the appeal, Dyson LJ, giving the judgment of the court, admitted that there was during the trial ‘*a plain commission of the prosecutor’s fallacy*’ (at para 20) and also that there were: ‘*shortcomings in the evidence and summing-up which were not made good*’ (at para 21). The Court of Appeal decided that the ratios were so large as to not fall foul of the guidance in *Doheny & Adams*, which discussed that the higher the ratios, the less impact the prosecutor’s fallacy would have on the proceedings.

4.12 R v George, 2007:

(Probabilistic reasoning errors)

The case of *R v George*, related to the firearm murder of the well-known television personality Jill Dando, on the doorstep of her home. Firearm discharge residue (‘FDR’) was found in the cartridge case and in the victim’s hair. The Defendant was one of a number of local suspects interviewed by the Police, and his coat was analysed by the FSS, who discovered a single particle (11.5 microns or the equivalent of about one hundredth of a millimetre) of FDR in the internal right pocket of the coat. The particle matched the constituent elements of FDR found in the cartridge case and on the victim’s hair.

At first instance trial:

In the course of the judgment, the court considered the FDR evidence, and stated that the two prosecution expert’s (Mr Keeley and Dr Renshaw) opinions were that the fact that only one particle of FDR was found was not significant. However, the defence expert’s (Dr Lloyd) opinion was that the single particle could have come from casual contamination by the police, due to poor evidence gathering procedures, and that to rely on it after one year of the killing was ‘incredible.’ The Court concluded that the FDR evidence was capable of supporting the

Prosecution case; and that its weight was a matter for the jury. The jury convicted the Defendant by a majority of 10:1.

After the trial, Dr Ian Evett, who from 1996 to 2002 worked for the Forensic Science Service ('FSS') and who had been developing an evidence analysis technique called Case Assessment and Interpretation ('CAI') - which adopts BT as its mathematical analysis tool, invited Mr Keeley, the prosecution expert, to apply the CAI technique to the FDR evidence on two different propositions: (1) that the appellant was the man who shot Jill Dando; (2) that the appellant was not the man who shot Jill Dando. Mr Keeley estimated that the likelihood of his finding no FDR particle had been 99 in 100 *on either proposition*, the likelihood of his finding one or a few particles as 1 in 100 *on either proposition* and the likelihood of his finding lots of particles as 1 in 10,000 (these figures being intended simply to signify 'remote in the extreme') *on either proposition* (at para 17). The significance of this was that, in Mr Keeley's opinion, the finding that he made of a single particle had in fact been 'neutral', which he then confirmed to Dr Evett.

On appeal:

The Court of Appeal commented that: *'It is often the case that a piece of evidence that proves nothing when viewed in isolation acquires probative relevance when considered in the context of other evidence'* (at para 31) and went on to say: *'We believe that the manner in which Mr Keeley gave his evidence was likely to have left the jury with the impression that his view was that, because you could discount each possibility of innocent contamination as 'most unlikely', the likely source of the particle was the gun that had killed Miss Dando: that was the only unexplored possibility'* (at para 39). In discussing the trial judge's summing-up to the jury: *'He went on to add that tests carried out under the supervision of Mr Keeley showed that FDR would more often be found on an individual firing a gun than not ...It is clear from these extracts that the summing up that the jury were directed that the evidence of Mr Keeley and Dr Renshaw provided significant support for the prosecution's case that the appellant had fired the gun that killed Miss Dando'* (at para 51).

The Court of Appeal stated that the FDR evidence did not make up the foundation of the prosecution's case and that there was other circumstantial evidence capable of implicating the

Appellant. However, since it would (now) be impossible to determine what weight the jury placed on the FDR evidence, the conviction was quashed and a retrial was later ordered.

At retrial:

At retrial, the trial judge excluded the FDR evidence as a consequence of the report made by Mr Keeley and Dr Evett, and the prosecution case was run on the basis of the eyewitness identification evidence, the police interview, the false alibi and bad character evidence – BG had a previous conviction at the Old Bailey for attempted rape, as well as various assumptions of false identities among other misdemeanours. However, arguably without the vital FDR evidence, the jury acquitted him.

4.13 People v Puckett (USA), 2008:

(Probabilistic reasoning errors)

At trial:

The case of *Puckett* was concerned with the 1975 rape and murder of Diana Sylvester. Sperm samples from the victim were taken and stored, since DNA testing did not exist at that time. An eyewitness gave the identity of a man she had seen at the time and place of the crime as ‘white, medium height, heavy-set, chubby, curly brown hair, beard, moustache, with a clean-cut appearance.’ After the arrest of Robert Baker, a local artist who seemed to fit the description, but who was released later without charge for lack of evidence, the case went cold.

In 2003, the police started taking DNA evidence from cold case files and matching them with databases of known criminals. The sample from Sylvester had degraded, so that only five whole loci and parts of two or three more could be matched. One candidate was matched: John Puckett, a 72 year old man with three counts of rape on his record from 1977. A photograph from the 1970’s was produced, which showed distinct similarities with the eyewitness’s identification, although Puckett was not arrested at the time of the crime.

During the trial preparation, Puckett's attorney, following an Arizona state-funded research project on DNA 'cold hit' matches, found that in a database of 65,000 profiles there were 122 pairs with 9 matching loci and 20 pairs with 10 – which should statistically be expected to occur in 1 in 13 billion people in the population, not 20 in 65,000.

The fact that people are surprised that there can be so many such matches reveals a fundamental misunderstanding of probability that is also known to be the case in the similar 'birthday problem', whereby for there to be an approximate 50/50 chance of any pair of people in a group having an unspecified matching birthday, you only need 23 people in that source group. What is easily overlooked when 1/13 billion is stated, is the fact that it is 'pairs' which is being measured, not 'individuals'. In the birthday problem, there may be 23 people, but there are 253 pairs of people, and thus the chance of finding a common birthday is greater than 50/50.

The argument that Puckett was a victim of the birthday problem, was rejected on the basis that in a 'cold hit' case, there is a specified profile with which to match the database profiles – the birthday problem simply looks for a pair of unspecified matching profiles.

Puckett's attorney then raised an argument relating to the meaning of finding 5.5 matching loci with the victim's killer and stating that this is a 1/1.1 million RMP. In Puckett's case, the database consisted of 338,000 registered sex offenders from California, which his attorney stated as meaning that he had an approximate 1/3 chance of being found as a match – not 1/1.1 million. The judge excluded this theory, but allowed the 1/1.1m figure to be presented to the jury, along with the eyewitness identification and Puckett's previous convictions. In fact what the lawyer should have said is that *'assuming the real attacker was NOT in the database of offenders, then there was still a 1/3 chance of finding a person matching the profile in that database'*. However, Puckett was convicted and sentenced to life imprisonment.

4.14 R v T, 2010:

(Probabilistic reasoning rejected by the court)

At first instance trial:

In *R v T*, the evidence in question was footwear marks, and the issue at trial was that of the identification of the murder suspect. The Defendant was convicted and appealed on the basis that the expert, Mr Ryder of the Forensic Science Service ('FSS'), had presented evidence unreliably on the likelihood of the footwear marks left at the scene of the incident matching those of footwear owned by the Defendant. Mr Ryder had extensive experience of footwear marks, having worked as a chemist at the FSS since 1989. His conclusion, as was written in his report and consistent later with his evidence-in-chief, was that there is a: '*moderate degree of scientific support for the view that the [Nike trainers] made those marks*' (at para 41).

Mr Ryder had arrived at his conclusion by adopting the recognised four-part formula for analysing footwear marks: (i) sole pattern, (ii) size, (iii) wear, and (iv) damage to the footwear. Where it cannot be said with any certainty that a footwear mark definitely did or did not match the suspect's, the term 'could match' would be broken down into sub-classes by the use of a likelihood ratio, which was adopted by various forensic scientists to promote consistency in judgements. It was the likelihood ratio of 10-100 which provided Mr Ryder with 'moderate support' for his evidential submission (at either extreme ends of the scale, <1-100 gives a verbal equivalent of 'weak or limited support, and >1,000,000 gives 'extremely strong support').

In court, Mr Ryder explained that he used BT to combine the four analysis points and arrived at his conclusion. His analysis of the sole pattern ('P') was supported by the FSS database comprising 0.00006 of all shoes sold in the year. Since the shoe size ('C') was judged to be 11, the Shoe and Allied Trade Research Association provided information which stated that size 11 shoes occur in 3% of the population. For wear ('W'), Mr Ryder concluded that half of the shoes could be discounted, and for damage ('D'), he concluded that almost no further shoes could be discounted which had not already been. After combining the figures $P(5) \times C(10) \times W(2) \times D(<1)$, he arrived at an overall likelihood ratio of ~ 100 . He did not submit his methodology in his report to the court. On cross-examination, Mr Ryder admitted that the defence lawyers had provided him with some information with which to calculate P.

On appeal:

After the first instance trial, the Appellant's counsel became aware that likelihood ratios were used in Mr Ryder's calculations and lodged an appeal on the basis that the Court of Appeal had already rejected the use of BT for all non-DNA evidence in the cases: *R v Adams (Denis)*, *R v Adams (Denis) No.2* and *R v Doheny*. The court analysed the international position on the use of BT and, using evidence provided by Mr Baldwin, the Forensic Science Regulator, concluded that in the Netherlands, Slovenia and Switzerland, there was evidence that BT is used, but not the extent. In New Zealand, BT is used for DNA, glass, paint and fibres. In the USA, experts are not allowed to use conditioning information (such as motive and opportunity) unlike in England, where this information is said to be a key component in calculating ratios, and, as a result, do not use likelihood ratios to determine footwear mark evidence. In Australia, many examiners used the USA approach, but some were looking at ways of developing an approach based on using likelihood ratio calculations.

The court rejected the Appellant's broad submission that the verbal scale should not be elucidated any further than the statement that the Defendant '*could have*' left the footwear mark. Instead, the court allowed future experts to make a definitive evaluative opinion on footwear marks, and although the court was not drawn to listing an exhaustive list of factors, it suggested, for instance, cases where there was an unusual size or pattern.

The court recognised that likelihood ratios have been used and accepted for DNA evidence in the domestic courts in *R v Doheny*, and in the New Zealand courts in *Lapper v R* and in Australia in *R v Karger*, *R v GK* and *R v Berry and Wenitong*, and stated that there was no authority presented to it where BT had been used for evidence with no '*proper statistical base*.' The court rejected an academic viewpoint that hard data was unnecessary to calculate likelihood ratios and that: '*... all probabilities are subjective and based on a combination of personal experience and the available data*' (at para 79), stating instead, that: '*An approach based on mathematical calculations is only as good as the reliability of the data used*' (at para 80) and that in this case none of the four-part footwear mark analysis could be said to have been determined by '*certain*' sources, for various reasons, including the small size of the database (8,122 shoes, where 42 million are sold every year), counterfeit shoes and Mr Ryder's own subjective uncertainty concerning wear and damage. The court went on to say: '*It is important to appreciate that the data on footwear distribution and use is quite unlike DNA. A person's DNA does not change and a solid statistical base has been developed which enable accurate figures to be produced*' (at para 83).

A brief mention was given to whether BT evidence should be presented to a jury. The court referred to the case of *R v Adams*, where Rose LJ explained that: *‘whilst Bayes theorem might be an appropriate tool for statisticians to establish a mathematical assessment of probability... It was inappropriate for use in jury trials for a number of reasons; the jury’s task was to assess the evidence by common sense and their knowledge of the world and not by reference to a formula’* (at para 89). From Rose LJ’s judgment, the court concluded that: *‘It is quite clear therefore that outside the field of DNA (and possibly other areas where there is a firm statistical base), this court has made it clear that Bayes theorem and likelihood ratios should not be used’* (at para 90). The court did, however, recognise that BT was used to analyse firearm discharge residue in *R v George*, but that the court in that case used the evidence to form part of the background to the court’s consideration of the overall evidence in the case and did not discuss the merits of the approach or if it was consistent with *R v Adams*, *R v Adams No.2* or *R v Doheny* (at para 90).

The court concluded that in an appropriate case, where a footwear mark expert were to give a more definitive evaluative opinion than *‘could have made’*, that no BT or other mathematical model should be used and, in any case, the word *‘scientific’* should be omitted as it may mislead the jury. Instead, the report should explain that its subjective findings are based on the expert’s experience. Therefore, due to the lack of transparency in how Mr Ryder’s evidence was presented to the jury, the appeal was allowed and the conviction quashed.

4.15 C v The Queen, 2011:

(Prosecutor’s fallacy)

At first instance trial:

In the rape trial of C, where a DNA sample was discovered on the victim’s clothing, the judge’s summing-up included the statement: *‘...the prospect of it being someone other than him by chance, about one in a billion’* (at para 45).

On appeal:

The Court of Appeal DID consider to the judge's summing-up comment to be an example of the prosecutor's fallacy. However, the appeal court refused to hold that the conviction was unsafe.

The reasons given by Leveson LJ were that the judge had given a: '*largely correct direction*;' and that in line with the judgment in *R v Adams & Doheny*, in which Philips LJ had stated that: '*...the more remote the random occurrence ratio, the less significant will be the adoption of the 'prosecutor's fallacy', until the point is reached where that fallacy does not significantly misrepresent the import of the DNA evidence*' (at para 46); that in *C* the proportions of the ratio were high enough ('*not less than 25 times the 1 in 40,000,000, as in Doheny*') to justify not rendering the conviction unsafe.

4.16 Conclusion:

As can be seen with this list of cases, there are some pressing issues which need immediate attention, discussion and research. Among other things, some key issues needing further discussion and which are subject to research in this thesis are:

- That committal of the prosecutor's fallacy is a routine problem and needs to be formally eradicated. Worryingly, the UK courts have now recently started to accept the committal of the prosecutor's fallacy. This problem has been examined in Chapter 3.
- That the UK courts have reached irrational and illogical conclusions as to which types of evidence are and are not capable of mathematical modelling. This will be examined in Chapter 5.
- That, prior probabilities - which lie at the heart of all Bayesian models - have not been adequately and comprehensively formalised, discussed or researched. This will be examined in Chapter 6.

- That, discrepancies in the beliefs of the role of the likelihood ratio can lead to incorrect probative value being attached to evidence. This will be examined in Chapter 7.

In summary, this thesis will deal with each of these problems, making recommendations for reform and, in doing so; seek to reduce the risk of miscarriages of justice due to mistakes in evaluation and presentation of criminal evidence. Each of these issues, along with the discussion of the pre-trial evaluation of evidence by the CPS, examined in Chapter 8, is a crucial step in supporting the research hypothesis.

5 CHAPTER V: Bayes and the Law

5.1 Introduction:

Following the examination of the evolution of courtroom mathematics in the last chapter's Digest of Legal Cases, this chapter examines the reasons why the UK courts have arrived at their current position of banning the use of Bayes' theorem (BT) in all but the narrowest circumstances, not only in the pre-trial evaluation of the weight of evidence by experts, but also in the way evidence is presented in-trial to, and evaluated by, juries.

The chapter begins by examining the background to the controversy, and explains the seminal 1971 anti-Bayesian stance taken by US writer Laurence Tribe, which has barely been eroded by the UK courts over the past forty-plus years. Tribe's main arguments are dissected, discussed and largely refuted within this chapter.

The chapter then examines the fundamental differences between explicit quantitative means of probabilistic reasoning as an aid to decision-making, such as BT, against other means, such as Wigmorean charts – a method advocated for use by a section of the mathematical community, supported by the Nuffield Foundation.

For the first time in any academic study, a parallel is drawn between Tribe's 1971 stance and the Nuffield Foundation's advice against the widespread use of BT and the UK courts' rulings. In addition, and also for the first time, the arguments within Tribe's oft-cited paper are broken down into its smallest composite parts and are systematically shown to be poorly reasoned, illogical and outdated, despite the UK courts' dogged insistence in supporting those general principles within Court of Appeal decisions, such as *R v Doheny & Adams* [1] and, more recently, *R v T* [2].

5.2 Background:

In criminal trials, the relationship between the fact finder – either a jury or a judge or magistrate sitting alone – and a lay witness should be one of healthy scepticism. While the witness purports to tell the truth of what he saw/heard/found at the time of the alleged crime,

it is the juror's task to ascribe weight to the evidence and having heard the entire case, decide whether the weight of the combined evidence is enough to return a verdict of guilt.

However, the relationship between the juror and an expert witness – especially where complex probabilistic reasoning is involved - while ideally no different to that of the relationship between the juror and the lay witness - is actually more one of trust and confidence in the expert's skill and experience; or as Annabelle Jones [24] puts it: *'juries clearly attach much weight to the views of learned experts... where the evidence involves probability, juries may be likely to follow the lead of the experts'*. The expert may have little or no first-hand knowledge of the instant case, but he or she has a wealth of knowledge in dealing with similar types of evidence, which can be used to devastating inculpatory or exculpatory effect for the defendant.

The juror, who also has no first-hand knowledge of the instant case and no means of collecting evidence himself, and is likely to have little to no experience in dealing with the field of evidence relevant to the trial, in his normal day-to-day life, maybe faced with complex and technical expert reasoning, which despite efforts to simplify the language, may mean that, at the heart of the matter, he must trust that the science behind the technical detail is sound. Law Commission Report *'Expert evidence in criminal proceedings in England and Wales'* [25] acknowledges the problem: *'...a jury, comprised as it is of lay persons, may not be properly equipped in terms of education or experience to be able to address the reliability of technical or complex expert opinion evidence, particularly evidence of a scientific nature. This being the case, there is a real danger that juries may simply defer to the opinion of the specialist who has been called to provide expert evidence, or that juries may focus on perceived pointers to reliability (such as the expert's demeanour or professional status)'*.

One type of expert evidence which has periodically pervaded the criminal trial process, concerns the use of mathematical reasoning to underpin the probative value of all of the other evidence in the case. As previously mentioned in Chapter 3, the advent of DNA profiling, where the evidence is presented as random match probability (RMP), has seen the use of explicit probabilistic reasoning increase exponentially as DNA evidence has become more widely used.

However, due to routine fundamental and sometimes basic reasoning errors in the probabilistic reasoning process by scientific experts, the national and international legal community have largely frowned upon, and have largely banned, the use of probabilistic reasoning in court. Cases such as *People v Collins* [5] in the U.S. and *R v Doheny & Adams* [1] and *R v T* [2] in the U.K. have evolved an almost total ban on the use of maths in court, and, arguably due to some poor reasoning and application of foundation principles, for good reason.

The mathematical community, who is largely convinced that probabilistic reasoning is an invaluable tool in calculating the true probative value of evidence, has been singularly unsuccessful in convincing many of the key members of the legal community of the same. The reluctance of the courts to encourage jurors to use a mathematical tool as an aid in their decision-making process stems from an overly complex application of probability theory, which has been said to ‘*deflect the jury from its proper task*’ [1]; and additionally, the use of probability theory has been banned from expert evaluation of types of evidence said not to derive from a ‘*firm statistical base*’ [2], due to its supposed lack of credibility. The court’s acceptance of DNA as firm, as opposed to ‘footwear marks’ etc which are supposedly not, has caused much consternation among the scientific community who believe that DNA is no firmer than any other type of evidence due to the inherently subjective nature of all probabilistic reasoning, regardless of the type of evidence the reasoning process is connected with.

The scientific community’s motivation in persisting to attempt to convince the courts to allow probabilistic reasoning in legal trials comes from its insistence that probabilistic reasoning is rife throughout the criminal justice system, albeit implicitly; so to ensure that the reasoning already in existence is transparent and accurate, a formal application of this reasoning is vital to ensure that trials are free of errors and fallacies. There is much credence in this stance, since there is evidence that fallacies are commonplace in legal reasoning, from the basic fallacy of the transposed conditional, or ‘*prosecutor’s fallacy*’ [15], to more sophisticated errors in reasoning due to mistakes and misunderstandings of the fundamental principles of probability theory. It is the scientific community’s stance that these errors can never be effectively eradicated by banning probability theory altogether, or by educating the players in a criminal trial, and so a mathematical model should be used as a tool to increase transparency.

Worryingly, because of the general and systematic misuse of probability theory by experts, there is also the danger that expert evidence has been presented to the jury with inaccurate and misleading calculations as to its probative value. Again, a mathematical model will expose these errors and fallacies before the evidence even reaches the jury, thereby reducing miscarriages of justice and saving valuable time and expense of numerous and lengthy trials and appeals.

Bayes' theorem, $(P(H|E) = [P(E|H) \times P(H)] / P(E))$, is one such evaluative probability method, designed to show altered certainty in a given hypothesis after observing evidence, and, as such, could be a good fit with the criminal trial process. However, since the correct application of this deceptively simple theorem is something that the scientific community have had little success in reaching a consensus over, it is not surprising that the legal community have shown a marked reluctance in adopting the theorem as a blanket decision-making aid. This is a pity, because in recent years and major advances in probabilistic reason software, the claims that the application to juries of Bayes' theorem is overly complex is now without foundation, much in the same way that the modern calculator - as has been suggested by Fenton & Neil [3] - has rendered complex mathematical calculations accessible to all.

However, although the solution to the problem of simplifying the calculations has been met by modern computers, the emphasis now is on ensuring that the input values placed into the mathematical causal model (see Chapter 2) are correct and reflect the correct reasoning process. This aspect of probabilistic theory can never be entirely placed out of human hands, as the criminal trial is a human process, and the mathematical model is still only a means of aiding that process. Mathematical modelling is not designed to usurp the juror's role, whatever critics of courtroom mathematics may think or say. One of the most vehement and enduring criticisms arose as a result of a prosecuting attorney's poor attempts at proving guilt by probabilistic means in the 1968 Collins case. The critic, Laurence Tribe, an advisor to the California Supreme Court, left a legacy of unresolved issues when he wrote his seminal 1971 article '*Trial by mathematics*' [4] as a rebuttal to Finkelstein and Fairley's [23] '*A Bayesian approach to identification evidence*'.

5.3 The early objections to the use of Bayes' theorem in court:

Following the ruling in Collins, Finkelstein and Fairley argued that since statistics will rarely conclusively identify a defendant, that a mathematical model is appropriate where statistical evidence alone is inconclusive. When other incriminating evidence raises a suspicion, apart from the statistical evidence; Bayes Theorem (BT) can be applied to indicate the degree that the inconclusive statistical evidence heightens the suspicion.

The 1968 Los Angeles case of Collins concerned a bag robbery perpetrated, according to an eyewitness, by a pair of robbers – a white woman, who did the bag-snatching, and a black man with a beard and moustache who drove their yellow Lincoln getaway car. At trial, the prosecuting attorney, Ray Sinetar, had a local mathematics professor, Daniel Martinez, give evidence to the jury that non-independent '*characteristic individual probabilities*' could be multiplied together to give a combined probability figure, using the product rule.

After multiplying these probabilities together and reaching a figure of 1/12,000,000 for a random match – which is to say that the observed trait evidence had massive prosecution probative value - the attorney assured the jury that it could be sure beyond reasonable doubt that the Collins' were the handbag robbers. The jury convicted the Collins' and they were sentenced to prison.

Lawrence Tribe, a Harvard mathematician, clerked for a Supreme Court judge as Collins' appeal eventually reached California's highest appellate court. Tribe advised the panel that Sinetar's submissions had no basis in sound mathematical reasoning. The Supreme Court found that B and E (see sub-section 4.2 above) could not be seen as non-independent probabilities, since the likelihood of having a beard is increased in men who wear moustaches. Also, the probability attributed to A was not made with any statistical support, and was merely the opinion of a mathematician with no experience of such things. Comment was also passed on whether C could ever be a true probability, on the basis that ponytails could be done or undone at will. F implied that the two people in the car were a 'couple', although their actual relationship was unsubstantiated.

In any event, even if the probability figure of 1/12,000,000 could be substantiated, this still does not mean that the Defendants were guilty beyond reasonable doubt – the number would simply represent a random match probability of coinciding events – therefore, this was a committal of the prosecutor’s fallacy. On this reasoning, the convictions were quashed. While the court ruled that the prosecutor’s mathematical reasoning was shoddy, it did not rule out the employment of probabilistic methods in future cases: *‘While we discern no inherent incompatibility between the disciplines of law and mathematics and intend no general disapproval or disparagement of the latter as an auxiliary in the fact-finding processes of the former, we cannot uphold the technique employed in the instant case.’* However, the court also held that *‘...no mathematical equation can prove beyond a reasonable doubt... that only one couple possessing those distinctive characteristics could be found in the entire Los Angeles area’*. Finkelstein and Fairley suggested that this ruling was conceptually ‘incorrect’, because the court was dealing with an existing finite population, the frequency with which couples with the identifying characteristics may be found in that population is identical to the probability of selecting one at random. *‘Thus, the court’s assumption that one in twelve million is a fair estimate of the probability of selecting such a couple at random necessarily implies that it is a fair estimate of the number of such couples in the population.’*

Presciently, the combining of prosecution and hypotheses of a *‘coincidental match in the world population’* is precisely the mathematical formulation needed to arrive at a random match probability (RMP) – the currently preferred measure of the probative value (for a definition and discussion of ‘probative value’, see Chapter 2) of trace evidence, such as DNA. The foundation for this reasoning is based on the principle, discussed by the Supreme Court in *Collins*, that in criminal trials the notion that a piece of evidence can be attributed to a defendant with 100% certainty is irrational: *‘The Collins court was right when it concluded that efforts to prove uniqueness usually will be futile. Few, if any, evidentiary traces can be demonstrated by statistical analysis to be unique to a defendant. Let us suppose a woman’s body is found in a ditch in an urban area. There is evidence that the deceased had a violent quarrel with her boyfriend, who has a history of violence, the night before. Investigators find the murder weapon, a knife which has on the handle a palm print which experts say appear in one case in a thousand. We now ask the significance of this finding.’*

Finkelstein and Fairley then explained that the prior probability (the estimate of the strength of a hypothesis *before* evidence is presented) is ‘critical’: *‘The statistical problem of the*

Collins case is that of estimating the very figure which the court took as its assumption, namely the probability that a couple selected at random would have the characteristics of the accused. That probability represents the frequency of couples meeting the description of the one placed at the crime. If a sufficiently precise estimate could be made that the frequency of such couples in the Los Angeles area was 1/12,000,000, it would be possible to state within reasonable margins for error that there was only one such couple in the Los Angeles area.' Therefore, by combining the prior probability by the RMP, a posterior probability can be calculated – this posterior reflects our belief in the hypotheses *after* evidence is presented, and it is this posterior which tells us of our measure of uncertainty of guilt. This uncertainty, Finkelstein and Fairley say, CAN - depending on the strength of the evidence in certain cases - be reconciled with the criminal standard of proof.

Laurence Tribe, in seeking to refute several of Finkelstein and Fairley's assertions, stated that: *'...mathematical devices may distort or destroy important values that a society means to pursue through its legal trials'* (although he conceded that in light of the Supreme Court's ruling that: *'there is no inherent incompatibility between the disciplines of law and mathematics; that if experts can be cross-examined and rebutted, the judge gives cautionary instructions to the jury and there is advance notice of the intent to use, coupled with a publicly financed expert to protect the accused, then there might seem to be no valid remaining objection to probabilistic proof'*). The article sought to discuss questions as to the use of mathematical tools in the course of a trial; whether parties should be allowed to use 'explicitly statistical evidence'; whether judges and juries should be allowed to use these tools as an aid to the decision-making process; whether the use of these tools was 'desirable' to the trial system as a whole and whether there was some means of quantifying the standard of proof set at 'beyond all reasonable doubt'. Tribe then went on to list the four fundamental mathematical errors made by the prosecutor in the Collins case as:

- (i) that the model was devoid of empirical evidence to support the values given;
- (ii) that in any event, that the product rule should not have been used to multiply the dependent factors, thereby giving a far too high probability figure;
- (iii) that in any event, the model did not take into account mistakes or lies of the prosecution witnesses; and

(iv) that the submission was a classic case of the prosecutor's fallacy [15], whereby the 1/12m random match probability was given as a probability of innocence, which should actually have been $\frac{1}{2}$, in light of the 24,000,000 population pool .

Tribe explained that there were three non-discrete types of evidence available for quantification: (i) *Occurrence* – whether proof that a certain number of a type of past acts not directly connected with the instant case can have any bearing on the instant case (this type of 'proof' was sought by Dr Meadows, the prosecution witness in the 1990's case of *R v Sally Clark* [9] who used national statistics of prevalence of sudden infant death syndrome (SIDS) deaths to infer a probability that the defendant murdered her two children instead); (ii) *Identity* - the main issue in the Collins case, which was as to the identity of the defendant(s); and (iii) *Intention* –whether the evidence that 1 in 20 fires are innocently caused can have any bearing on an arson case.

Tribe then conceived of a hypothetical 'Blue Bus' case to illustrate how not only could 'objective frequency not translate to subjective beliefs', but also how for policy reasons, it would be unfair to expect a party who may be statistically guilty for a majority of claims to be held responsible for all of them. In the Blue Bus case, Tribe suggests to consider '*a case where a plaintiff is negligently run down by a blue bus. The question is whether the bus belonged to the Defendant. The plaintiff is prepared to prove that the Defendant operates four-fifths of all the blue buses in town. What effect, if any, should such proof be given? The upshot would be a regime in which the company owning four-fifths of the blue buses, however careful, would have to pay for five-fifths of all unexplained blue bus accidents – a result as inefficient as it is unfair.*' This hypothetical seeks to highlight the problems in reconciling the criminal standard of proof of 'beyond all reasonable doubt' (BARD) with probabilities.

In essence, the proposition made here is that since 'reasonableness' is unquantifiable, that the criminal justice system, possibly for reasons of policy, has sought to create a standard of proof which does not carry a permitted margin of error. This argument promotes the idea that to allow an acceptable margin of error would expressly acknowledge a weakness in the system, which could be exploited. Also, any quantified doubt, should by rights, be attributed to the defendant, since '*society does not tolerate the conviction of some innocent suspects in favour of the vastly larger number of guilty criminals*'.

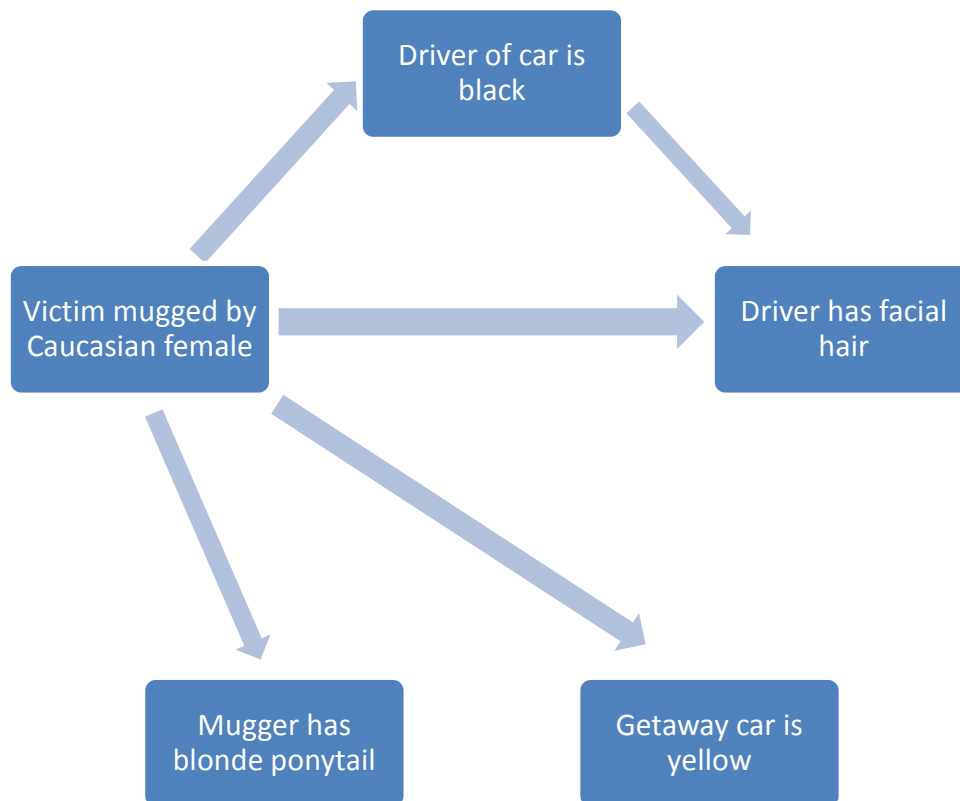
Tribe also argued that probabilities should be the preserve of possible future events, not criminal trials where the alleged event has certainly happened, which of course is a basic fallacy, as discussed in Chapter 2. Nevertheless, at one stage, he concedes that BT, as Finkelstein and Fairley points out, *could* provide a solution to combining ‘non-mathematical’ (presumably eye witness statements, exhibits, etc) with ‘mathematical’ evidence (such as blood types and, in the modern day, DNA, etc).

However, this application of a probability theorem is not without its risks, namely: (a) the difficulty in assessing an accurate prior on the non-mathematical evidence; (b) avoiding an overpoweringly large prior which will overshadow the other evidence in the case; (c) the problem of ‘soft variables’ – the argument that some probabilities cannot be counted, which therefore means that the variable does not exist and therefore cannot be inserted into the model; (d) the lack of concentration on the right types of evidence, such as mental state or motive, which would not be treated by the model; (e) the problems in double counting conflicting or dependent evidence. Since legal cases are complex affairs, with evidence arising at inconsistent times due to examination and cross examination of witnesses, and revelations occurring at unexpected junctures, the application of BT would have to necessarily reflect this complexity. Tribe suggests that the necessary adjustment to BT to encompass the enmeshing of the evidence is thus:

$$P(X|E1+E2) = P(E2|X+E1) / P(E2|E1) \times P(X|E1)$$

This would mean that the theorem could not be applied sequentially as the evidence is presented, but would have to be applied in the ‘*terribly cumbrous*’ form shown above. Of course, we now know that Tribe’s model here is an ill-defined three-node Bayesian network (BN) where E1 and E2 are parents of X. A properly defined BN would eradicate risks of dependence and circular reasoning where more than one piece of evidence is presented. In fact, Edwards [12] presented a BN for the Collins case as long ago as 1991 as follows:

Fig 5.3



Edwards explains the structure of the Bayes net as asserting ‘*non-independence between colour and facial hair. But it assumes conditional independence between skin colour and car colour, and between skin colour and facial hair of the male and hairstyle of the female. These conditional independence assertions could be refuted by evidence, requiring addition of arcs to the net. One may also potentially need more nodes. If the defence had cast doubt on the credibility of the eyewitness, that testimony would require addition of one or more nodes*’. However, even with the potential for adding nodes, what can be seen here is that the BN is not only a simple graphical representation of what could be considered a complex case with many pieces of evidence, but it helps the jury understand the potential for dependencies and ‘double counting’ of certain evidence leading to unfair weight being attached to a case.

On other policy considerations: Tribe argued that a mathematical model would bring about an end to the presumption of innocence, since the juror would have to arrive at a prior of guilt before evidence is presented. And as far as ‘*dehumanisation of justice*’ is concerned; Tribe make the statement that the jury’s job is partly ritualistic – that a mathematical model can never replace the need for humans to be part of process which settles conflicts peacefully according to the needs of society.

The final part of Tribe's article critiques a proposed new way of quantifying the criminal standard of proof. The 'Jury Decision Model', created by Kaplan and Cullison, suggests a method for aiding a jury's decision to convict by replacing the current 'beyond reasonable doubt' standard with the concept of 'utility', whereby the juror weighs up the balance between acquitting a guilty person and convicting an innocent one.

By comparing preferences between convictions or acquittals for what the juror considers a 'guilty' or 'innocent person' (depending on the facts of the particular case), the juror creates two scales: (i) 'conviction of a guilty person' (most desirable) to 'conviction of an innocent person' (least desirable), and compares this first scale with: (ii) 'acquittal of a guilty person' (least desirable) with 'acquittal of an innocent person' (most desirable) – the parameters of each scale set between 0 and 1.. The method is that the evidence in the case will lead to a degree of comfort for the fact finder to place a point in each scale. For example, if the case concerns a brutal murder, the degree of risk to society of freeing the defendant is high – this may weigh a decision towards conviction where the prior belief is 50/50.

Tribe argued that in some cases where the juror is only, say, 75% convinced of the defendant's guilt, this would lead to a conviction, which falls far short of the current BARD standard, and would be undesirable. Tribe also argued that decisions are not so easily scaled between the two sets of parameters given by Kaplan and Cullison, since if a crime is considered by a juror as 'unimportant', he may set the preferences to convict at 0, regardless of the evidence.

Tribe's main objections to the use of probabilistic reasoning in court are numerous, and have received either criticism or support from academics and the legal community regularly and consistently until the present day. The main areas of contention, still largely unresolved, can be summarised as follows:

- (i) That an accurate and/or non-overpowering prior cannot be devised.
- (ii) That not all evidence can be considered or valued in probabilistic terms.
- (iii) That some evidence cannot be mathematically valued and therefore be inserted into the model, due to the existence of 'soft variables'.
- (iv) That no probability value can ever be reconciled with BARD.

- (v) That due to the complexity of cases and non-sequential nature of evidence presentation, any application of BT would be too cumbersome for a jury to use effectively and efficiently.
- (vi) That probabilistic reasoning is not compatible with the law, for policy reasons.

The interesting thing to note is that regardless of whether there is any credence to Tribe's claims – and mathematical academics have been refuting them for years – the mathematical community has never weakened the legal community's stance on allowing mathematical reasoning in court. In fact, the ruling in the fairly recent case of *R v T* [2] has shown that the judiciary's distrust of probabilistic reasoning has moved from a ban on the overt use of BT by juries and other fact finders during the trial process, to banning the behind-the-scenes use of BT by experts, while in the solitude of their own laboratories, when arriving at quantified values for presentation of evidence to juries. But where did this seeming revulsion derive from? If each of Tribe's main objections are examined, and the academic response to these objections are equally measured, it may shed some light on the issue.

5.3.1 (i) That an accurate and/or non-overpowering prior cannot be devised:

Finkelstein and Fairley recognised that the formulation of an accurate prior probability was crucial to the integrity of the probabilistic model in the Collins case, which sought to prove a 'certain match' between the crime scene trace (in that case an eyewitness statement of the suspect's physical characteristics) and the defendants' own physical characteristics, as they asserted that the '*approach in Collins makes the number of suspects critical*'. Regardless of whether 'uniqueness' or 'traits' are sought to be proved by a mathematical model; what is essentially meant here is that if a model purports to provide a posterior based as a result of the reduction of a large pool of suspects, the original reference group must encompass the suspect but not be so large that the model loses fails to prove the point it was designed for. As they continue: '*Setting a generous upper bound will usually defeat the proof*'.

Therefore, the problem is that a fine balance needs to be struck between introducing too many suspects into the reference class, and too few. It was stated that this decision was wholly 'arbitrary', since it was difficult to make an objective assessment as to what would be a sufficient geographical area to draw the pool of suspects from. Ward Edwards [12], in discussing the best way to formulate the correct reference class, and therefore achieve a

logical prior, in Collins, explains: *‘The prior probability in this case would be the probability that a random man/woman couple who could have been at the time/place of the crime and who might have an interest in committing it would be guilty. In this instance, 10,000 (which seems very high) couples might exist, giving a 0.0001 prior.’* As can be seen here, the prior is based on a guess of the upper boundary of how suspects there are likely to be, to prove the given hypothesis (in the Collins case, at least) that the suspect is the person(s) seen by the eyewitness(s) at the scene of the alleged crime.

Richard Friedman [26] argues that in cases where identification of the suspect is a main issue, and there is trait evidence linking the suspect to the alleged crime scene, that Finkelstein and Fairley’s approach of using a reference class is not only correct but is consistent with trial practice. He recognises that there MAY be some credence to Tribe’s argument that a reference class prior means that the model starts with a value larger than zero – which means that at the outset of the case, the defendant already has a degree of guilt attributed to him before any evidence has been observed – a fact which sits uneasily with those who believe that all defendants should commence their case with the presumption of innocence valuing their guilt at 0.

However, the problem with a prior of 0, is that all evidence conditioning the prior results in a posterior of 0 (as Chapter 2 has shown). A result that is not possible, according to those who believe that all probability is subjective (and therefore subject to uncertainty to whatever degree). Friedman calls this class of doubters ‘Bayesioskeptics’, which arguably pitches two camps: those who believe in Bayes theorem, and those that don’t – a distinction which is not accurate, since (a) the pro- and anti-Bayes supporters may agree on some aspects of probability theory and not others; and (b) that it implies that there is such a thing as a ‘Bayesian’ – purportedly a cultish group who have different values, ideas and principles to those ‘who do not believe’.

This artificial distinction is not helped by Richard Friedman: *‘Application of the conventional probability theory to subjective probability assignments is the essence of Bayesianism’*; or by Ward Edwards: *‘Bayesians consider probabilities to be an appropriate metric for any uncertainties about an empirical proposition on which the outcome of a decision depends, and consider all contexts in which decisions are made to be appropriate for their use.’* Even the Nuffield Foundation [16], one of whose aims is to promote better relations between the

legal and mathematical community, state: *‘Although the Court of Appeal has denounced attempts to encourage jurors to attempt Bayesian calculations, especially in relation to non-scientific evidence, many forensic scientists are confirmed or unconscious Bayesians and routinely employ likelihood ratios in the course of generating expert evidence ultimately adduced in court.’* This sort of distinction can do nothing to create a collaborative and friction-free working environment. In fact, there are even some academics [27] in the mathematical community who advocate the use of a ‘Bayesian approach’ to probabilistic reasoning, without recommending the ‘full application of Bayes’ theorem’. This confusing stance must be hard to follow for those non-mathematicians who are unaware that the distinction between Bayesian calculations and Bayes’ theorem is artificial at best, and misleading at worst.

Friedman’s problem with the current formula for generating priors in identification cases is summed-up as follows: *‘Bayesioskeptics have contended that the presumption of innocence is inconsistent with a non-zero prior probability of guilt. A zero prior probability of guilt represents a belief in innocence that cannot be shaken no matter how strong the evidence. I believe this argument is wrong, but I do not believe the easy answer that has been offered by Bayesians is quite satisfactory. The easy answer in a criminal case is to say that the presumption means that, before any evidence is presented, the fact-finder should treat the defendant as no more likely as anyone else, or, in a variant, as equally likely as anyone else, to be guilty. I have at least four problems with this test:*

- (i) It is indeterminate: there is no obvious single reference population.*
- (ii) The defendant may contend that he is substantially less likely than the ‘mean’ member of the population to have committed the crime.*
- (iii) Information that the fact-finder may legitimately bring into court, without the need for evidence, may suggest that the defendant is more likely than most other members of the population to have committed the crime; consider, for example, the outset of a case in which the jury knows nothing but that the charge is armed robbery and that the name of the defendant is John (suggesting that he is male and therefore more likely to have committed armed robbery than the average member of the population).*

Friedman’s argument is therefore in support of BT, inasmuch as the ‘Bayesioskeptics’ stance that ‘a model is impossible to create’ is wrong; yet he says that the ‘Bayesian’ formulation of a prior from reference class populations is also wrong. Friedman has not recognised that the

prior probability is based upon the witness's (whether expert or lay) own combination of hypotheses. After all, the prior must be based on some evidence, even if it is evidence of the world-at-large, rather than evidence of the case itself.

The mathematical community have argued that the prior is immaterial, as it will be personal to the fact finder and not open to scrutiny – after all, the use of BT is designed to illustrate the impact of evidence on priors, not to scrutinise the prior itself. Ronald Allen [28] has a problem with this approach though: *'Bayes' theorem requires the formation of a prior probability of any proposition to be updated. How that is to be done within the law has proven elusive. Moreover, the purely subjective view that any prior is as good as any other leads to radical indeterminacy and has no obvious connection to accurate adjudication. A non-subjective view suffers from the lack of any relevant data. Before hearing the evidence, how would any juror have any non-subjective basis to form a probability of, say, O.J. Simpson's guilt?'* Allen's argument is therefore that subjective probability is incompatible with criminal trials, and objective probability (or assigning values to the fact finder; the very anathema of Bayesian reasoning) is pointless, due to its lack of personal reflection to the task at hand. In other words, Allen's view is that Friedman is wrong – probabilistic reasoning and the law are wholly incompatible, regardless of the methodology.

Friedman suggests that BT can be used by a juror merely to 'check' on his own reasoning process: *'The rule expresses the commonsense proposition that if a piece of evidence is more likely to arise given an hypothesis than given its negation, the evidence makes the hypothesis more likely. In the second model, an observer might at any time make a set of probability assessments that are inconsistent. If the inconsistency is a glaring one, the cognitive dissonance will be apparent, and the observer will adjust the assessments to bring them more closely into line. This adjustment might involve altering any probability assessment, including a prior probability.'* Johan Bring [29] argues that Friedman's suggestion that BT could be used merely as a tool for elucidating reasoning, without a requirement of formulating a prior pre-trial, is illogical: *'Friedman does not seem to be too strict about the requirement that the prior should be assessed before hearing the evidence. He suggests that if a fact-finder is unhappy with her posterior she might go back and change, for example, her prior. So, if the result from the Bayesian calculation does not correspond with our intuitive belief we could alter some components in the Bayesian calculations until the Bayesian answer corresponds with our intuitive belief! What then, is the point of Bayes' theorem?'*

Bring's stance is that if BT should be used to provide answers as to the probative value of the evidence in a case, the fact finder should not be allowed to simply adjust the model to suit his beliefs – since the model should elucidate beliefs.

There is much credibility in this argument – how would a fact finder know if the intuitive belief he has is a product of a fallacy, if BT doesn't illustrate it for him? However, Friedman could simply be seen here to be side-stepping any difficulties surrounding the formulation of an accurate prior, by advocating the mathematical model as a means to quantify, visualise and aid the juror's reasoning process. This objective towards transparency must surely be supported by the legal community.

To sum-up, the general consensus seems to be that those sceptical of the use of probabilistic reasoning in legal trials have used the controversy surrounding the prior as a lynchpin in entirely dismissing the approach as inappropriate. The supporting camp must do better in showing that the prior probability can be formulated without the model providing inaccurate values – the very thing that the model is designed to eradicate from the fact finder's normal (non-probabilistic) reasoning process.

Up until this point, the argument is unresolved. Tribe's position is that a prior cannot be formulated, as is Allen's and Bring's. Friedman states that a prior can be formulated, but the current use of reference classes in cases where the probative value of identity evidence is being calculated, is flawed. In effect, although the argument has become more sophisticated and nuanced, we are no further on in this matter than Tribe's original 1971 objection. Therefore, the problem with priors will have to be fully addressed later in Chapter 6.

5.3.2 (ii) That not all evidence can be considered or valued in probabilistic terms:

The Nuffield Foundation, in its most recent practitioner guide [30], accepts that there is no such thing as individualised evidence: '*Individualisation*'. *This type of opinion relates to the issue of the individualised source of recovered material, or trace, and is usually an expression of a personal conviction. Examples of individualisation include:*

- *This shoe made the mark at the scene*

- *The piece of broken button recovered from the complainant's bed came from the broken button that remains on the jacket of the suspect*
- *The fingermark is that of the left thumb of the suspect.'*

A key feature about evidence in a legal trial – rather than evidence in, say, medical testing – is that the witness is not the fact finder, as Nuffield recognises: *'Even though an opinion of personal conviction may well be based on vast experience and proven competence, they are very difficult, if not impossible, to justify on strictly logical grounds. To arrive at an opinion of an individualisation, the expert would need to be sure that it was impossible to obtain the observations if any other alternative proposition were true...Events that are impossible have probability zero. However, in situations like the above, zero probabilities are very difficult to justify. The probabilities may indeed be very small; approaching zero, but it would be incorrect to say 'It is impossible'. In these situations, the opinion of an individualisation is a personal, albeit possibly highly expert, conviction.'* In medical testing, the witness (normally the doctor or other expert) views the testing procedure and the results firsthand, whereas in criminal proceedings, the fact finder must NOT be someone connected with the case. Therefore, every piece of evidence comes with some doubt as to its provenance, necessitating some kind of mathematical model to chart the uncertainty. (Interestingly, there could be some similarity between medical and legal cases in that the true fact finder in medical testing could be seen to be the patient him/herself, and the doctor as the witness – with the patient making the decision whether or not to accept the doctor's findings and embark on the treatment offered – which means that regardless of the discipline, the uncertainty in evidence is always inherent and present.)

Finkelstein and Fairley [23] recognised that it is rarely likely that a fact finder will believe 100% the evidence observed by him: *'The Collins court was right when it concluded that efforts to prove uniqueness usually will be futile. Few, if any, evidentiary traces can be demonstrated by statistical analysis to be unique to a defendant.'* Of course, Finkelstein and Fairley were underplaying their assertion – in fact, there will never be a situation where a fact finder has rationally eradicated ALL of his uncertainty in a piece of evidence. Therefore, if one accepts that all evidence carries an element of uncertainty, it MUST follow that evidence is probabilistic, and is somehow capable of being represented by an ascertained tangible value less than 1.

One of the fundamental difficulties, for those not familiar with using quantified numerical values to elucidate reasoning normally associated with non-quantified statements of belief – such as ‘I’m pretty sure it will rain today’, is that there is tendency to think that numbers are too precise and do not adequately deal with issues of uncertainty in the same way that words do. ‘Pretty sure’ is a familiar term to most people, but ‘58% sure’ is only within a few people’s comfort zones. The supporters of BT are comfortable converting ‘pretty sure’ into a percentage, but non-supporters state that this conversion can be misleading and in some ways impossible to make.

This debate has led to two distinct and opposing camps forming. The supporting camp believe that all evidence is capable of being converted to numerical values in preparation for insertion into a model, while the non-supporting camp believe that quantified values can never fully convert into a fact finder’s true beliefs due to an inherent unquantifiable element present in some kinds of evidence. What makes some kinds of evidence ‘unquantifiable’, and not others, is open to debate, but the issue is not helped by ill-defined criteria for what makes some evidence reliable, and therefore quantifiable. The Court of Appeal in the case of *R v T* [2] certainly have been led by some members of the mathematical community, on this point, to believe that the distinction exists, but this is by no means a universal, or perhaps even majority view.

Perhaps, one way of delineating the difference between ‘reliable’ (and therefore quantifiable and capable of model insertion) and ‘unreliable’ evidence, is to ascertain whether the reliable evidence has any inherent qualities. Finkelstein and Fairley examined a distinction between what they termed ‘statistical’ and ‘non-statistical’ evidence by stating: *‘The court reversed the Collins’ conviction because it felt that the powerful statistics would fool a jury into overlooking the possibility that the basis for the calculations could be in error. The court was obviously right. However, correct statistical methods will usually have an effect opposite to that feared by the Collins court. Findings based on such statistics should generally weaken non-quantitative testimony based on the same evidence. Statistical observation is of attributes that can be objectively measured; it cannot hope to have the richness of information involved in ordinary or educated recognition.’* With this approach, Finkelstein and Fairley are advocating that quantified evidence IS too precise and because of this, is less convincing to a fact finder.

Ward Edwards [12] disagrees with this stance: *‘One major argument against the use of probabilities in court is that probabilities are inherently and inappropriately precise because they are numbers. Any empirical scientist will assert that virtually any number derived from observations is imprecise. A careful scientist reporting an empirically derived number, such as a person’s height, will report it as $y \pm z$, where z is a number measured in the same units as y reflecting measurement error, typically a standard deviation or a half-range. Physical measurements typically lead to values of z much smaller than y ; a report of height might be 6 ± 0.5 feet; the error is normally counted from repeat measurements, carried out by the user or manufacturer of the measuring equipment. Apart from measurement errors, errors can also occur in counts where you cannot be sure you have counted all cases you set out to count, or where a count depends on categorization and there is a borderline case. The point is that no empirically derived number is either inherently precise or imprecise. How precise it is depends on its nature and the process that links it to observations. In this respect numbers do not differ from words.’* As can be seen with this argument, Edwards is actually arguing that uncertainty arises in all walks of life – whether data is witnessed first or second hand. In fact, Edwards is arguing that all evidence is ‘subjective’, which is to say that there must at some point in the witness’s mind a ‘leap of faith’ made between what is empirically shown and what he believes – in other words:

1. A metre stick can only ever be measured against other metre sticks, and who is to say that they are correct either?
2. Even in the case of a single metre stick, the minute discrepancies which may arise between the object and the observer, and other observers, also need quantifying rather than ignoring.

Edwards, explains: *‘The type of evidence typically called ‘statistical’ is based on measurements obtained empirically. No-one doubts the relevance of such evidence, though Tribe does not support the view that numbers are the natural way to report it. The argument is about its sufficiency. All of evidence about identity is of this kind. Tribe seems to be saying that if such evidence is inherently insufficient, then identity is not establishable in court. In Tribe’s ‘blue bus problem’ how do we know that the errant bus was blue, or even that it was a bus? To modify the hypothetical, a reliable colour blind witness’s testimony that the vehicle was bus is conclusive, but he can say nothing about its colour. But buses have distinctive number plates and he wrote down the result. To reproduce Tribe’s hypothetical; we could specify that blue buses and only blue buses have 123 as the first three digits of their license*

plates, and that the witness saw those three digits, but not the last three. To specify a condition that surely would meet Tribe's requirement for individualised information, we could have the witness report the full license number 123456, which is indeed the license number of a blue bus... Where does 'statistical' evidence stop and some other kind begin? The answer to this question must be arbitrary.' By showing a fundamental flaw at the very heart of Tribe's Blue Bus case, Edwards displays a healthy scepticism for any data which is said to be empirically sound.

Therefore, if all evidence is uncertain, it is better to be upfront and honest about the uncertainty by modelling it, rather than simply banning a model on the basis that some uncertainty within it exists, since as Tillers and Gottfried [31] state: *'Probability theory assumes the existence of uncertainty. (One might say that probability theory takes uncertainty seriously.) The entire point of using probability theory is to talk coherently about uncertainty—not to eliminate uncertainty.'* Edwards' argument renders the Blue Bus case invalid and it exposes Tribe's deeply fundamentally flawed probabilistic reasoning, as he seems to support the notion that some evidence can be taken as fact.

It is perhaps this foundation thinking that separates the subjectivists from the 'frequentists' (those who believe that empirically derived information is the only useful tool in valuing data). The legal community, as evidenced in cases such as *R v T* [2], seemingly reject this subjectivist approach to probability. Richard Friedman [26] also recognises Tribe's flaw: *'As regards the 'blue bus hypothetical' (that the Blue Bus Company owns 80 per cent of the buses in town. Should the Blue Bus Company be liable?): The 80 per cent figure is not a probability, but only a datum. Probability assessments are made by observers on the basis of all the data received -which include information about what is not received. Sometimes the data received are so crisp and strong that they narrowly confine the range of probability assessments that most rational observers could make, but that is not this case. Failure of either side to produce more evidence might lead a rational fact-finder to assess a probability substantially higher or lower than the 80 per cent figure.'* Therefore, once evidence is presented to the fact finder, it will be his task to assign a probability value of his own. In fact, Friedman's argument exposes another flaw in Tribe's reasoning: that Tribe has obviously disregarded that it is the fact finder's role to assign his own subjective value to evidence, not the witness's - regardless of whether the evidence is quantified or not.

Tillers and Gottfried [31] champion subjectivist probability by arguing that while some probability may be grounded in empirically gathered ('reliable') data, some is not: *'Probabilities are not the same thing as statistically grounded probabilities. Yes, modern statistical analysis does involve the probability calculus. But, as the word 'statistics' implies, statistical analysis involves and requires systematic collection of data or observations, data and observations that can be summarized in the form of statistics. It is possible to talk—and talk coherently—about odds or probabilities without systematically gathering data, compiling statistics or analysing systematically gathered collections of data. In short, although it is not possible to do statistics without doing probability, it is possible to do probability without doing statistics.'* The important thing to note is the authors are suggesting that probabilistic reasoning is not to be confused with the gathering of, and inferences drawn from, 'reliable' information.

The court in *R v T* seemingly have disregarded this distinction when ruling that probabilistic reasoning must only be adopted to draw inferences from data drawn from *'a firm statistical base'*. This distinction between probability drawn from 'reliable' information, and probability drawn from what can only be termed 'unreliable' information may be part of the reason why the Court of Appeal in *R v T* felt that subjectivist probability has no place in the courtroom. The authors use of the words 'systematic collection of data and observations' implies that if the negative were true, that there would be no system in place – in effect cutting to the very heart of the problem; as the legal community are plainly rigidly committed to the creation and maintenance of tangible systems.

It is possible that making the distinction between statistics and probability has led to confusion with the distinction between 'reliable' and 'unreliable' information. In the Nuffield Practitioner Guide No.1 [16], a current published guide to the use of probabilistic reasoning in legal trials – purportedly bridging the divide between the legal and mathematical communities – the distinction is made thus: *'Statistics are concerned with the collection and summary of empirical data. Probability is a branch of mathematics which aims to conceptualise uncertainty and render it tractable to decision-making. Hence, the field of probability may be thought of as one significant branch of the broader topic of "reasoning under uncertainty".'* This supposed endorsement of the use of probabilistic reasoning for decision-making would do little to dispel the legal community's fears that non-statistical information is reliable. In fact, the idea of 'empirical information', which means 'systematic

collection and recording', would likely lead any casual observer to believe that subjectivists simply 'pluck figures out of the air' – an idea which is unsurprisingly unpalatable within the legal community.

The Nuffield Guide No.1 makes the distinction between subjective and objective probabilities: *'Probability can be "objective" (a logical measure of chance, where everyone would be expected to agree to the value of the relevant probability) or "subjective", in the sense that it measures the strength of a person's belief in a particular proposition...It is not always possible to obtain a good estimate for a population relative frequency based on sample data: relevant datasets may be incomplete or non-existent. In these circumstances, relative frequencies may be replaced by estimates based on an expert's personal experience and knowledge of the type of evidence in question.'* Again, it is the distinction between 'good estimates' which are based on sample data, rather than simply 'estimates' based on an expert's personal experience and knowledge which may be another factor in the Court of Appeal's decision to ban BT in court proceedings.

In addition, this stance is completely at odds with Robertson and Vignaux's [32] seminal 1998 recommendation that there should, in principle, be no difference between the way we evaluate traditionally quantified evidence (such as DNA) and other types: *'While DNA evidence may be highly "statistical", other scientific evidence, such as fibre analysis, is less so. Nor is there any reason in principle why less quantitative evidence such as eye-witness identification should not be subjected to statistical analysis.'* Biedermann et al [33] also support the view that subjectivist probability is compatible with the legal trial system: *'...the subjectivist view appears to be the most adequate approach for addressing an individual's situation of making judgements under uncertainty. Accordingly, probability assignment in individualisation processes is not a case of disputing facts, but a case of declaring probabilities that sincerely represent the experts' actual states of belief.'* It is again arguably the lack of cohesion between members of the mathematical community which fosters such little confidence in it when it attempts to promote its ethos to exterior factions such as the legal community.

One thing that can, and has been, agreed upon by the mathematical community, is that probability must be used to ascertain the value of evidence alleged by the prosecution to have derived from the crime scene. This so-called 'trace evidence' can never be attributed with

100% certainty to any defendant. The reason for this is that because certain physical traits proliferate around the world population and are shared by some members - and the means of judging 'matches' is based on matching certain traits within the sample, with the defendant's own traits.

Therefore, whether the matching process is based on objective probability (so-called 'empirically measured' data) – or subjective probability (estimates based on experience and knowledge) – the end result is that all trait evidence carries a degree of uncertainty and this uncertainty carries a tangible value. This view was expounded by the Nuffield Foundation in their Practitioner Guide No.1: [16] *'The ensuing decades have witnessed a growing realisation that all scientific evidence is probabilistic and no current forensic technology supports unique identification of individuals. DNA is different only insofar as it wears its probability on its sleeve, whereas other sciences and technologies have tended to conceal their probabilistic foundations in ostensibly binary concepts such as 'match'/'no match'.'* In Nuffield Practitioner Guide No.2 [19], DNA testing is placed into a special category of sciences which expressly uses probability as a tool to display the probative value of a crime scene trace: *'One of the most distinctive features of DNA profiling, as compared with older and hitherto more established branches of forensic science and forensic medicine, is that DNA evidence is explicitly probabilistic. An expert witness does not – or at any rate, certainly should not – identify a particular individual as the donor of the genetic material from which a DNA profile was produced. This is because the standard DNA profile is produced from only a small sample of the donor's entire DNA. Thus, even if DNA itself is assumed to be unique to each individual, more than one person could still share the same DNA profile, e.g. more than one person could be 'a match' to crime scene DNA... In fact, DNA profiling will never be able to produce a verifiably unique match to a particular individual, because the evaluation of DNA evidence is always, in part, a question of probability.'* This means that since in Guide No.1, the Nuffield Foundation state that all scientific evidence is probabilistic, this means that all evidence is probabilistic – what makes evidence 'scientific', as they suggest, must be arbitrary. Even fingerprint matching, once the 'acme of forensic identification' [34] is now seen as probabilistic.

Fenton and Neil [3] state that where there is probability, there is a natural propensity towards making fallacies – they recommend Bayes' theorem as a means by which the risk from miscarriages of justice, due to fallacious reasoning, can be reduced through education in the

workings of Bayes' theorem: *'There is unanimity among exponents of BT, that a basic understanding of Bayesian probability is the key to avoiding probabilistic fallacies.'*

There is, however, an alternative to express quantification of probabilistic evidence, which may provide an answer to those who are sceptical of using Bayes' theorem in the courtroom. The Wigmorean method uses a similar causal model to illustrate the relationship between pieces of evidence in a complex case – but without quantifying the nodes - thereby circumventing the argument that some evidence is incapable of quantification. What the Wigmorean method DOES purportedly do, is help the fact finder formalise his reasoning process.

5.3.2.1 The Wigmorean method:

The Nuffield Foundation, in its third guide [35], describes the Wigmorean method as: *'...nothing more (or less) than an attempt to summarise the logic of inferential reasoning in graphical form, tailored to specific intellectual (analytic and decision) tasks. It is, in other words, a practical heuristic for litigation support designed specifically to assist those who need to formulate, evaluate or respond to arguments inferring factual conclusions from mixed masses of evidence to improve the quality of their intellectual output,'* and the key difference between the Wigmore method and Bayes' theorem thus: *'Bayes nets share superficial similarities with Wigmore charts, in that both employ graphical symbols to represent inferential relations in a visually vivid and useful form. However, there are important differences between them. Bayes nets are expressly probabilistic and can be used to calculate the strength of an inference as well as mapping logical relations between propositions.'* Therefore, while the Wigmorean method is still causal in nature, it does not seek to elucidate what the inherent probabilities represented by the causal nodes actually are.

The Wigmorean method has been used for many years. Kadane and Schum [36] used the method to chart the evidence in the controversial Sacco and Vanzetti case, which to this day, has academics debating the flaws in the reasoning which may have led to a miscarriage of justice for the now-executed defendants. *'Once produced, the charts resemble the Bayes' nets and directed acyclic graphs which have been developed more recently by those interested in the analysis of evidence, a fact which suggests just how far ahead of his time Wigmore was.'*

Once the evidence has been charted the authors move beyond Wigmore by undertaking a probabilistic analysis of the evidence. Here their favoured tool is Bayes' theorem which is used to produce likelihood ratios for the evidence in the various inferential chains leading to a conclusion.' As can be seen in Mike Redmayne's above critique of Kadane's use of the charts, he accepts that although the Wigmorean method is ostensibly used as means of charting causal links in evidential chains, it is not until Bayes' theorem is used to formulate likelihood ratios does the model become 'probabilistic'.

5.3.2.2 The key difference between the Wigmorean method and Bayes' theorem

In fact, it could be said that the main difference between Wigmore and Bayes is that the lack of likelihood ratios means that the Wigmorean method carries no possibility of formalising or checking the accuracy of the pairs of hypotheses which make up a likelihood ratio. Both methods rely on the formulation of pairs of hypothesis, but only Bayes theorem provides a rigid framework for ensuring that the hypotheses pairings are accurate. If express values are not given, how can the fact finder be absolutely sure that the witness (whether expert or lay) and himself are 'speaking the same language'? If an eyewitness states that he is 'fairly certain' that the defendant is the suspect, how can the fact finder compare this statement with other evidence from the same eyewitness that he is 'almost sure' that the co-defendant is not the suspect? Surely if percentage values were expressly given for each statement, the fact finder would be able to make a clearer calculation of the probative value of both pieces of evidence?

The Nuffield Foundation HAS recognised this key limitation in the Wigmorean method: *'The logic of inferential reasoning depicted in a chart is objective and must conform to the dictates of rationality, but the charting exercise as a whole is subjective in this important sense: the quality and utility of any given chart and keylist turns crucially, not only on the amount and quality of material information available for analysis, but also on the analytical skill and imagination of the person constructing the chart. Thus, it is said that a chart is "a map of the [charter's] mind, rather than a map of the world". Different people will construct different charts from the same facts and evidence; and no two charts (probably, not even two charts produced by the same analyst at different times) would be entirely alike in every nuance or finer detail.'* [35] This potential disparity between the witness's and the fact finder's view of the case could be catastrophic for the wronged defendant.

While the supporters of Wigmore over Bayes acknowledge the importance of formalising the evidential reasoning process, in order to reduce potential miscarriages of justice, the furthest they are prepared to go to adopt formalisation is to have the witness ‘write down’ his thought processes. The flaw in this approach, as far as the Bayes’ supporters are concerned, is that there is no more precise language than the language of numbers - which means that a model WITHOUT numbers may be even more imprecise than no model at all.

This lack of quantitative reasoning – relying on words rather than numbers – is definitely problematic. As the Nuffield Foundation itself acknowledges: *‘One further significant limitation of the Wigmorean Chart Method is that it represents only the structure of inferential reasoning, and not the relative strength of particular inferential arguments or the probative value of particular pieces of evidence... The problem was that Wigmore’s symbols for “provisional”, “strong”, “doubtful” and “weak” probative force were simply reports of his own subjective impressions, with no standardised metric or internal logical structure.’* [35] This admission of the limitation of the Wigmorean method would likely lead an impartial reader into believing that the Nuffield Foundation is a strong advocate for quantitative reasoning for all evidence in legal trials.

However, as will be shown here, nothing could be further from the truth. In fact, it may even be argued that since the Foundation is so influential in its rarefied position of ‘bridge’ between the legal and mathematical communities that it is highly likely that the appellate courts in decisions such as *R v T* have been influenced by the Foundation’s lack of support for Bayes’ theorem in reaching its conclusion to ban the use of the theorem by not only juries but also by experts in the privacy of their own laboratories. The Foundation has shown a keen interest in the use of the Wigmorean method: *‘We can all still learn from Wigmore if we are prepared to listen to what he has to teach... The extraordinary flexibility of Wigmorean method merits emphasis. Anybody concerned with fact finding, fact analysis, or formulating, challenging or evaluating arguments about facts can easily adapt the method to their particular requirements. All standpoints and professional roles can be catered for, within or outside the legal process.’* [35] While no-one could reasonably object to the recommendation of any visual aids to decision-making, including the Wigmorean method, there is an underlying problem that may potentially go unnoticed; that the method could be promoted in favour of other quantitative methods.

The first indication of the Foundation's feelings towards the suitability of some types of evidence for quantitative evaluation can be seen here (this comes from the Foundation's first practitioner guide): *'It should be borne in mind, however, that although most evidence adduced in criminal proceedings does not come with a pre-assigned quantified numerical value attached (e.g. what is the probability that an eyewitness identification is accurate? Or the probability that a confession is true?), much forensic science evidence (including DNA profiling) is predicated on quantified probabilities and is consequently directly amenable to Bayesian calculations'* [16]. The statement that DNA is 'directly amenable to Bayesian calculations', arguably shows some support for an argument that DNA is somehow more reliable than eyewitness identifications and confessions.

In fact, the Foundation seems to show an innate distrust for representing probabilities in a range, regardless of whether the evidence is DNA or another type, such as with confidence intervals, in any legal proceedings: *'However, confidence intervals and related judgements of statistical significance are not appropriate measures of the value of evidence in criminal proceedings, for several important reasons. First, the selection of a confidence level is subjective and arbitrary. Why 95%? Why not 99% or 99.9%, or for that matter 75% or 70%? Levels of confidence which are conventionally regarded as satisfactory in social science research have no bearing on the level of confidence ideally required for epistemically warranted verdicts in criminal proceedings.'* The use of the words 'epistemically warranted verdicts in criminal proceedings' shows that the Foundation has no faith in things that they say cannot be 'known'.

By implication, the argument that confidence intervals are not compatible with knowledge shows the Foundation's lack of confidence in probability theory as a whole. Confidence intervals are designed to capture knowledge, yet the Foundation seems to use this aim *against* the science. Fenton and Neil [3] argue that confidence intervals should not be the preferred method of reducing uncertainty anyway. Their approach is to use a full Bayesian approach based on subjective prior assumptions about the sample space. They argue that any other way of approaching uncertainty will lead to a *'contrived way that confidence intervals are generated and interpreted'*, due to their fundamental basis being founded on wrong and irrational assumptions concerning the uniformity of infinite sample spaces and sampling techniques within them.

The Foundation then argues that where something is not ‘known’, it would be best not to give quantitative values in case these values are misconstrued as ‘fact’: *‘There is the worry that impressionistic judgements of inferential strength, once concretised in charted symbols, may assume a solidity they scarcely warrant, becoming de facto fixed points in the chart impervious to reconsideration and potentially skewing further analysis. All in all, it seems best to keep the chart(s) and keylist free of impressionist judgements of the strength of inferences and the probative value of evidence.’* [35]. This is a barely-concealed attack on the use of probabilistic reasoning. Since ‘impressionistic judgments’ means ‘subjective probability’, then this passage is clearly saying that they should not be used in court, for fear that a juror may attach too much weight to the value and not feel that the weight is subject to cross-examination. This principle must be at the very heart of the Court of Appeal’s reasoning in *R v T*.

Now that it has been shown that there are distinct similarities between the legal community’s and the Foundation’s negative views on the use of subjective probabilities in court, there can be also seen a certain support, from the Foundation, of the courts’ decisions in cases such as *R v Doheny & Adams* and *R v T* to not only ban probability theorem from legal proceedings, but to ensure that it never returns. One method of ensuring this is entrenching a principle of non-use of mathematical reasoning which will prove hard to dislodge. The Foundation explain the problem faced by the court in *Doheny & Adams*: *‘In the leading case of Adams (Denis), the Court of Appeal, twice, went out of its way to condemn any attempt to encourage jurors to employ formal mathematical models when evaluating evidence presented at trial. This is entirely consistent with orthodox legal theory stipulating that jurors should arrive at their verdicts using their ordinary common sense reasoning: it is precisely their ordinary common sense, untainted by specialist knowledge, which qualifies jurors as “expert” decision-makers on the common law model... A growing number of forensic scientists is already utilising Bayes nets as a way of understanding the meaning of the evidence for themselves, typically by modelling alternative possibilities consistent with analytical results, so that they can then pass on this better understanding to the police, prosecutors or defence lawyers instructing them... None of this implies that jurors in criminal trials need to know the first thing about Bayes nets. The Adams principle is entirely unaffected.’* The problem here is that there is no such thing as an ‘Adams principle’.

The case of Adams was heard in 1998 – and the case was presented with many errors in a highly complex way. There have been massive strides in advancing the science and creation of software to reduce complexity – this is the luddism which Roberts has said shouldn't exist: *'The point is to use litigation-support tools effectively, being mindful of their limitations, rather than to discard a tool simply because it has limitations – a luddite strategy which would result in throwing much more than Bayes Theorem into the fire.'* [35] The major problem here is that mathematical reasoning is somehow divorced from common sense – while mathematical reasoning may not be 'common', it cannot be said that it is not correlative with common sense.

As the Foundation reiterate: *'Fact-finding in criminal adjudication is, generally speaking, accomplished by ordinary common sense reasoning rather than through the application of mathematical formulae, as the Court of Appeal emphatically reiterated in Adams'*. [16] It is this two-fold support for the court's decision which makes it almost impossible for future mathematicians to make a case for BT: Firstly, that it is not common sense; and secondly that since it has been promoted badly in the past, that it should not be used in the future. Further, the Foundation sees no merit in trying to resolve these issues, by educating jurors in the use of probability theory: *'It would in theory be possible to teach jurors to calculate likelihood ratios in the same way that many forensic scientists currently do. This would, of course, be a major departure from traditional trial practice, and the Court of Appeal strongly deprecated any developments in this direction in R v Adams where the defence had attempted to instruct the jury in the use of Bayes' Theorem:*

[This indented passage is taken from the *R v Adams* judgment] *"[W]e regard the reliance on evidence of this kind... as a recipe for confusion, misunderstanding and misjudgement, possibly even among counsel, but very probably among judges and, as we conclude, almost certainly among jurors. It would seem to us that this was a case properly approached by the jury along conventional lines.... We do not consider that [juries] will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion. We are very clearly of opinion that in cases such as this, lacking special features absent here, expert evidence should*

not be admitted to induce juries to attach mathematical values to probabilities arising from non-scientific evidence adduced at the trial.”

The Nuffield Foundation continues: *‘Those of a Bayesian disposition might be tempted to interpret these remarks as a victory for the dark forces of ignorance over the light of science. This would be hasty and excessively pessimistic conclusion. Juries are empanelled in order to inject common sense reasoning into criminal adjudication. But this does not mean that criminal trials are a forensic free-for-all. Both the content of the information presented to juries, and the manner of its presentation, are carefully regulated by the law of criminal evidence and procedure. To this extent, the jury’s common sense reasoning is constrained and channelled in conformity with the rule of law.’* [19] This ‘support’ for the Court of Appeal’s judgment and pre-emptive strike on those who seek to disagree with it, is a clear statement of intent that BT should never be introduced back into the trial process.

The last sentence in this passage is very telling, and an extremely strong statement: *‘...the jury’s common sense reasoning is constrained and channelled in conformity with the rule of law.’* This implies that the jury’s use of BT would break the moral foundations of the legal system.

This argument that most types of evidence should not be quantitatively valued could not be more clearly put by the Foundation than in the following statement: *‘Judgements of probative value are fundamentally qualitative and only incidentally or secondarily quantitative. The testimony of a single independent and reliable eyewitness will often defeat five dodgy alibi statements; just as one compelling argument trumps fifty flimsy make-weights. The aggregated assessments of probative value required to determine whether a normative standard of evidential sufficiency has been satisfied, e.g. whether the prosecution has proved its case “beyond reasonable doubt” (or so that the fact-finder is “sure” of the accused’s guilt), must by extension be qualitative at their core.’* [35] What is being stated here is that even if DNA is expressly quantitative, that even that type of evidence will never be enough to convict a defendant.

This is precisely Tribe’s point made in 1971, and is essentially the Foundation’s position today, which means that there is no pro-probabilistic reasoning academic viewpoint in the last half-century which has interfered with this key principle.

In fact, the Crown Court Bench Book '*Directing the Jury*' [37] advises that in complex cases a written 'Route (or Steps) to Verdict', which are a '*logical series of questions couched in words which address the essential legal issues*', including a 'summary of evidence', should be provided by the trial judge to the jury, based on the following recommendation: '*The judge is not obliged to repeat every byway taken by the evidence, but is entitled to assess what is important and what is peripheral*'. The book provides an example of summing-up for a sample of DNA evidence: '*Mr B is able to say that the chance of finding another match with a person in the UK population unrelated to the defendant is 1 in 1 billion. The population of the UK is about 60 millions. It is for you to decide whether in the circumstances of this case that effectively excludes anyone else*'. The problem here is that the judge is effectively and inadvertently combining two pieces of evidence: (i) the DNA sample; and (ii) the reference class (or 'location evidence') that the defendant belongs to – the population of the UK – thereby advising the jury that the pieces of evidence have a bearing on each other. While this may or may not be true, there is no advice given as to HOW the evidence may be combined. In Chapter 6 of this thesis, it is demonstrated by Bayes' theorem that the traditional method of combining trace evidence with location evidence, adopted in cases such as *R v Doheny & Adams* [1], is wrong, yet the Bench Book provides no means of allowing the jury to autonomously formulate the correct means of combining evidence, thereby entrenching the error. The Route to Verdict in its unquantified form, and the Wigmorean method, are therefore similar in approach and equally flawed.

While the Foundation seems prima facie to support the use of BT to address some of the limitations of the Wigmorean method, with the following statement: '*Bayesian Networks (often shortened to "Bayes nets") are similar to Wigmore charts, in that they attempt to model inferential reasoning (including compound or catenated inferences – inferences upon inferences) through formal models represented graphically by a simple collection of symbols. The major difference is that Bayes nets calculate quantified probabilities for alternative propositions, that is to say, they purport to measure probabilistically the probative value of evidence or the strength of evidential support for particular arguments or entire legal cases. Bayes nets therefore answer directly to a significant limitation of Wigmorean method, i.e. the inability of Wigmore charts to provide much if any guidance on quantitative issues of weight, probative value or degrees of inferential strength needed to satisfy legal burdens of proof*' [35]. The problem is that because the Foundation has already stated that evidence should not be evaluated quantitatively, that this limitation on the Wigmorean method is immaterial. In

other words, while BT can potentially remedy the defect in the Wigmorean method, it cannot be used, which means that the probability of ever having probabilistic reasoning introduced into the legal trial system must be low.

It is perhaps now easy to see how some key members of the legal community have been reluctant to introduce probabilistic reasoning into the trial process. At the outset, a number of probabilistic reasoning mistakes (in evaluation AND presentation of evidence), in historic cases gestated the notion that law and probability are incompatible.

There were careful arguments for and against the introduction of a model which may redress some of these issues, but the crux of the problem, is that the Nuffield Foundation (ostensibly set up to decrease the divide between supporters and naysayers of probabilistic reasoning in court by providing simple, transparent information amenable to both legal and mathematical communities), has effectively supported the naysayers. This has happened in a logical process:

1. By suggesting that ‘some types’ of evidence are amenable to quantitative analysis.
2. By advocating the Wigmorean method of non-quantitative evidence analysis.
3. By inventing the ‘Adams Principle’; this seeks to entrench the idea that BT should never be used by juries.
4. That while acknowledging that BT is a useful tool in presenting some evidence, such as DNA, it should not be used for other types.
5. That since the court in *R v T* has delineated between these two types of evidence, that this effectively gives rise to the invention of the ‘*R v T Principle*’; another example of an attempt to entrench an idea that BT should not be used even as a means of presenting the probative value of evidence to juries.
6. By not seeking to educate juries or legal practitioners in the operation of BT; that this situation can never change unless outside help is found.

7. By not acknowledging that there have been many advances in science since the case of Adams; that the legal community would unlikely trust that the mistakes made in 1998 would not happen again.

In short, the court judgements and the findings of the Nuffield Foundation, an organisation trusted by the legal community, has effectively given weight to the argument that probabilistic reasoning should not be introduced at any stage of the legal trial process, unless it is used by expert witnesses to present the probative value of DNA and similar types of ‘reliable evidence’. Perhaps unsurprisingly, this is precisely the position in the British criminal justice system today.

Amazingly, The Nuffield Foundation [35] still believes that there is a fundamental difference between the uncertainty in an event that has not happened yet and one that already has: *‘Causation is another recurrent area of confusion. When interpreting Bayes nets, it is tempting and quite natural to think of one node exerting a causal influence on another, or others, at the end of the arc. However, the relationships represented by Bayes nets are explicitly characterised as probabilistic rather than causal relationships. Some probabilistic relationships are causal, but most are not. For example, the probability of its being 25 degrees Celsius at the weekend is causally related to the probability that I will get sunburnt when I go to the beach on Saturday. However, the probability that the perpetrator has blond hair has no causal relation whatsoever with the probability that Adam, who has blond hair, is the perpetrator. The vast majority of probabilistic relationships depicted by Bayes nets are of the second, non-causal, Adam’s blond hair variety, rather than variants of the first, causal, my sunburn kind.’* This kind of argument might lead to an assumption that Bayesian probabilistic reasoning has no place in criminal trials, since a trial concerns ‘non-causal’ relationships between the event and the defendant. Of course this stance is simply yet another way of saying that some things in life are objectively certain and some are not. Fenton and Neil [3] dispute this argument on the basis that: *‘Uncertainty is the same whether the events have happened or not and whether they are unknown or not’*. This must be the correct view.

5.3.3 (iii) That some evidence cannot be mathematically valued and therefore be inserted into the model, due to the existence of ‘soft variables’:

This argument rests not on the premise that evidence *should* be presented probabilistically, but whether it possibly *can*, as Allen [28] argues: *‘Even the simplistic version of Bayes’ theorem reproduced above requires the articulation of a likelihood ratio. In some cases involving relative frequencies or obvious situations, this may plausibly be formed, but many cases will not involve either relative frequencies or obvious situations.’* The capability of evidence to be presented probabilistically is a controversial subject because it cuts right to the core of the age-old issue of whether probability is objective or subjective in nature.

If probability is thought to be subjective, there would not be a database in the world big enough to convince the subjectivists that the information contained therein is accurate, since many errors can have crept in during collation, as Aitken and Taroni [38] recognise: *‘In some circumstances a negative (non-match) result may be reported where the two samples have a common source, and hence should provide a positive (match) result. This is a ‘false negative’: the result reported is negative, but that result is false. A ‘false positive’ report, conversely, occurs when a positive (match) result is reported, when the two samples in reality have different sources and therefore should have been reported as a non-match. False negative and false positive reports can arise for a host of reasons (the details of which need not concern us here) including contamination of samples, laboratory testing error and misinterpretation of test results.’* The problem with subjective probability is that it is ultimately uncertain between individuals. Since a ‘leap of faith’ is required for one person to believe that something is probabilistically so, this faith is unlikely to transmit well between individuals – in other words, what one may trust with his own eyes, the other does not necessarily trust unless he witnesses the same event.

However, if probability is thought to be objective, it asserts what a reasonable person would believe to be true. Objectivists rely on databases of information, which are taken to be reliable as the ‘best’ means of judging the truth of a proposition. The problem with objective probability is that it relies heavily on information which has been empirically tested, collated and stored, which means that information which has not been rigorously subjected to this treatment cannot become part of the database. It is this type of information, which has not been tested, which are called ‘soft variables’ – or as Tribe puts it: *‘...if you cannot count it, it does not exist’*. The crux of the problem, therefore, is: ‘Do soft variables exist?’

The Court of Appeal in the case of *R v T* believes that only DNA and other types of evidence said to derive from a '*firm statistical base*' are capable of probabilistic modelling; which means that it believes that DNA is empirically sound. Evidence such as footwear marks is not empirically sound because there is not a recognised national footwear mark database. There are, however, a number of experts who have experience in dealing with footwear mark evidence, but these people's professional views are not backed up by objectively sound databases, which means that their opinions are simply that; opinions.

These opinions should not be allowed to be transposed into numerically quantified values for fear that these values become 'too solid', as the Nuffield Foundation puts it: '*There is the worry that impressionistic judgements of inferential strength, once concretised in charted symbols, may assume a solidity they scarcely warrant, becoming de facto fixed points in the chart impervious to reconsideration and potentially skewing further analysis.*' Whether or not this 'worry' has any basis in truth, the fact remains that there is an obvious correlation between the Nuffield Foundation's reasoning and the Court of Appeal's views. In *R v T* the evidence in question was footwear marks, and the issue at trial was that of the identification of the murder suspect.

In court, Mr Ryder, the prosecution expert, explained that he used BT to combine the four analysis points of the footwear mark found at the crime scene and arrived at his conclusion. His analysis of the sole pattern ('P') was supported by the FSS database comprising 0.00006 of all shoes sold in the year. Since the shoe size ('C') was judged to be 11, the Shoe and Allied Trade Research Association provided information which stated that size 11 shoes occur in 3% of the population. For wear ('W'), Mr Ryder concluded that half of the shoes could be discounted, and for damage ('D'), he concluded that almost no further shoes could be discounted which had not already been. After combining the figures $P(5) \times C(10) \times W(2) \times D(<1)$, he arrived at an overall likelihood ratio of ~100. He did not submit his methodology in his report to the court.

On cross-examination, Mr Ryder admitted that the defence lawyers had provided him with some information with which to calculate P. '*It is important to appreciate that the data on footwear distribution and use is quite unlike DNA. A person's DNA does not change and a solid statistical base has been developed which enable accurate figures to be produce...It is quite clear therefore that outside the field of DNA (and possibly other areas where there is a*

firm statistical base), *this court has made it clear that Bayes theorem and likelihood ratios should not be used.*' As can be seen from this judgment, DNA has been elevated, through international recognition of its development as a 'certain' science, above that of more 'impressionistic' types of evidence. While Jackson et al [39] argue that '*... all probabilities are subjective and based on a combination of personal experience and the available data*', one of the factors which could be viewed as repugnant to the court is that Mr Ryder had used defence information to calculate P.

This means that not only did the footwear mark evidence values not come from an empirical source, it did not derive from the expert's own knowledge and experience either. Of course, it is a valid argument to suppose that asking other individuals opinions adds to the stock of communal knowledge and experience, but this information was arguably no more than casual conjecture – after all, what motivation could there possibly have been for the defence to have provided either (a) accurate values; or (b) values which were not deliberately skewed against the prosecution case?

The Nuffield Foundation [16] are even sceptical of empirically gathered information in databases, whatever the type of evidence: '*It must be stressed, however, that statistical inferences are ultimately only as good as their underlying data, which in turn depends upon (1) the appropriateness of the research design (including sampling methodology) and (2) the integrity of the processes and procedures employed in data collection.*' This scepticism is easily transmutable into blanket distrust for probabilistic reasoning in court, by a legal community who look to the mathematical community for guidance in such matters: '*Experts in particular fields may be willing and able to advise on the relative strengths and weaknesses of particular reference samples, or may operate with their own assumptions. Ultimately, however, it is for the legal system to determine whether such data adequately support particular inferences for the purposes of criminal adjudication.*' By opening the door to the potential problem in this way, not offering a solution, and then leaving it to a non-expert community to formulate their own opinion as to which evidence is appropriate for modelling, it is no wonder that the non-expert courts would feel safer rejecting the entire science rather than entertaining any future possibility of finding a middle ground.

Is not all knowledge and experience simply based on opinion, if not derived from firsthand knowledge of an event? After all, even a 'firm statistical database' is only as firm as the

opinion of the person creating it. Tillers and Gottfried [31] are very forthright on the argument that a database cannot exist which is large enough to eradicate all uncertainty: *‘The notion that soft variables cannot be quantified is a myth. For example, I can and do make uncertain judgements about how my neighbour will feel next time I see her—and, if asked, I can and will tell you what I think are the chances that I am right. There are, of course, more systematic critiques of the notion of the unquantifiability of soft variables, but it is not useful or possible to review them here. Suffice it to say that there is an entire family of approaches to probability—they can be called personal or subjective probability—that rest on the assumption that ALL (or almost all) probability judgements are expressions of personal and subjective probability estimates. Proponents of this sort of interpretation of the probability calculus maintain (roughly) that there are no ‘objective’ probability estimates or ‘hard’ variables that disclose their relevant statistical probabilistic properties entirely without the corrosive intervention of subjective human judgement.’* This is clear denial of the existence of hard or soft variables, yet the Nuffield Foundation [16] seem intent on taking the opposite view: *“Relevant “data” are of different types. Towards the harder end of the spectrum, experts may be able to draw on extensive surveys, databases or experimentation. At the softer end of the spectrum, the only available relevant data may be the expert’s personal experiences and memories of previous casework... Irrespective of their quality and status, data enables the expert to assign a likelihood (or probability) for particular findings that is necessarily personal and subjective, even in relation to ostensibly “hard” data.’* In effect, the Foundation’s argument can be summarised shortly; that subjective probability derives from a logical and legitimate school of thought, but it is not necessarily appropriate for use in legal trials. This is obviously a complete contradiction, and easily misinterpreted by the legal community.

In fact, Redmayne et al [40] support some aspects of the Court of Appeal’s decision in *R v T* - namely that the calculations used should have been made explicit - even if they reject the court’s main conclusions: *‘There is, in fact, much to welcome in the Court of Appeal’s judgment in R. v T, starting with the court’s commendable determination to subject the quality of expert evidence adduced in criminal litigation to searching scrutiny.’* Of course, this is a noble stance, since efficiency and transparency in the criminal trial process is key to a just system; but the underlying problem is that of whether this ‘searching scrutiny’ would uncover any less quality in the evidence of an expert using skill and experience, than one relying on a database that he had no control over creating? Further, even if he did create the

database, what level of scrutiny could there possibly be which would ever be enough to create certainty?

Even if many different individuals supported the empirical soundness of a database, this does not mean that the information contained therein is not subject to individual interpretation of results. Friedman [26] argues that *‘Different fact-finders may rationally make different probability assessments if for no other reason than that they approach a problem with different bases of information from the outside world. Thus, whatever the value of Bayesian methods as opposed to classical statistical methods in scientific inquiry, in litigation I believe that a subjectivist approach to probability is the only one that can offer any hope of assisting in the analysis of juridical proof.’* Of course, one answer is to say that if even DNA is uncertain due to uncertainty of the so-called ‘firm database’, then maybe probabilistic modelling should not be used for even THAT type of evidence.

The Nuffield Foundation [19] lists the various potential weaknesses in DNA profiling: *‘The probative value of a DNA profile, quite irrespective of its notional weight, hinges crucially on a series of prosaic assumptions, including the following: (i) genetic material from which a DNA profile could be generated remained at the crime scene, without irremediable degradation or contamination; (ii) the physical sample was collected properly at the crime scene (or from the suspect, victim, or whatever); (iii) the sample was successfully transported to the laboratory without interference or contamination; (iv) at the laboratory the sample was analysed using appropriately calibrated and properly functioning machinery, in accordance with appropriate scientific protocols; (v) the results of the tests were accurately observed and recorded; and (vi) at no stage during laboratory testing procedures did the sample become contaminated with other genetic material, wrongly labelled, switched with other samples, etc.’* However, simply because there are weaknesses in a science, this does not mean that the baby should be thrown out with the bathwater. Fenton et al [11] propose that the only way to deal with uncertainty is not to ignore it, but to model it.

The ‘event tree’ which pertains to the following discussion can be found in Chapter 2: *‘In order to demonstrate how a simple match evidence case quickly scales to become too complicated for intuitive comprehension, we present an analysis of a basic case of match evidence that includes negative and positive testing errors. When we allow for the possibility of testing errors, the following relevant information must be considered:*

- Prosecution hypothesis (H1): “The defendant is the source”
- Defence hypothesis (not H1): “The defendant is not the source”
- Evidence E1: “The source profile is tested to be of type X” (note: we can no longer assume the source profile actually is type X)
- Evidence E2: “The defendant profile is tested to be of type X (note: we can no longer assume the defendant profile actually is type X)

Because of the probability of false positives we cannot assume from the above evidence that either the source or the defendant have type X. Instead these assertions are also unknown hypotheses:

- Source type hypothesis (H2): “The source profile really is type X” (true or false)
- Defendant type hypothesis (H3): “The defendant profile really is type X” (true or false)

What we have, therefore, is a problem involving five ‘variables’ H1, H2, H3, E1, E2 which can all be true or false (in order to do the necessary Bayesian reasoning). But this means there are 32 different scenarios representing the different possible true/false combinations (although some are and some are not observed, such as the evidence being false, and some are logically ‘impossible’, such as the defendant is the source and the source is type X while the defendant is not type X).

NB. Scenarios for the prosecution likelihood include all scenarios that stem from the branch H1=true, which assumes the prosecution hypothesis (defendant is the source):

- Scenario 1 (this is the ‘normal’ prosecution scenario) in which H1, H2, H3, E1 and E2 are all true.
- Scenario 2 (this is an often ignored prosecution scenario) in which H1 is true (the defendant is the source) but the defendant is not actually type X. Both the test of the defendant and source, however, incorrectly result in an X.

NB. Scenarios for the defence likelihood include all scenarios that stem from the branch $H1=false$, which assumes the defence hypothesis (defendant is not the source):

- Scenario 3 (this is the ‘normal’ defence scenario) in which the tests are correct but the match is coincidental.*
- Scenario 4 this is the defence scenario in which the defendant is incorrectly tested to be type X.*
- Scenario 5 this is the defence scenario in which the source is incorrectly tested as type X.*
- Scenario 6 this is an often ignored defence scenario in which both the source and defendant are wrongly tested to be X.’*

With Fenton et al’s approach, the uncertainty inherent in even DNA testing – currently thought of by the Court of Appeal as a sufficiently ‘certain’ science (presumably meaning that it is not subject to scrutiny such as is proposed here) – is brought to the attention of the fact finder.

The problem with the Court of Appeal’s judgment in *R v Adams* and *R v T* is that this scrutiny will never be afforded DNA evidence, since Bayes’ theorem has been barred from use by juries. This is, as Redmayne et al [40] observes: ‘*hardly in keeping with the Court of Appeals ringing endorsement of transparency.*’

To sum up; Fenton et al’s argument that even DNA or similar types of ‘database’ evidence should be subjected to Bayesian evaluation, due to uncertainty inherent in ‘false positive’ matches, means that soft variables cannot exist. If they did, this would mean that the corollary, ‘hard variables’, also exist – which as can be seen here, cannot be true.

Therefore, the only logical answer is that all evidence carries uncertainty, but this uncertainty must be modelled in order that the fact finder is made aware of the weaknesses in presented evidence, and can deal with these weaknesses either by acknowledging them and reconciling personal beliefs with them, or by rejecting the witnesses’ accounts of his/her uncertainty. In this way, the fact finder’s role in the legal trial is preserved.

5.3.4 (iv) That no probability value can ever be reconciled with ‘Beyond all reasonable doubt (BARD)’:

The criminal justice standard of proof, set universally at ‘beyond all reasonable doubt’ (BARD) acknowledges that the trial system is inherently uncertain, as Laudan [41] states: *‘Although the juror must be ‘fully convinced’ of guilt, the standard is not set at 100% - a level of proof which no-one believes is available’*. Any system which relies on proving a hypothesis through evidence is uncertain, but in the criminal justice system this uncertainty is magnified due to the fact finder never being allowed to be a witness to any part of the event under scrutiny – this is why jurors are asked pre-trial what they know of the case at hand, in order that they must be precluded for having been tainted by any firsthand knowledge.

BARD acknowledges that allowing a small amount of uncertainty into the criminal trial evidence will not discourage a fact finder from returning a rightful guilty verdict; as the Nuffield Foundation [19] recognises: *‘There is no such thing as absolute, complete, unimpeachable and non-revisable certainty in the empirical world. Human decision-making, in other words, occurs under conditions of unavoidable uncertainty. This is clearly reflected in orthodox conceptions of the criminal standard of proof as ‘beyond reasonable doubt’, not beyond all doubt, or every conceivable doubt, etc.’* The value of ‘reasonableness’ is, in its very nature, unquantifiable. However, the problem is that any model which returns a value for guilt at below ‘1’, is expressly stating that the BARD standard has not been reached, which means that theoretically the model carries an inherent flaw.

Laudan explains that society will not accept any ‘explicit’ admission that the criminal justice system is flawed in this way: *‘To give a probability figure for the threshold would devalue the standard. Any specification of a degree of belief necessary for a finding of guilt (such as 95% confidence) involves an explicit admission that wrongful convictions will inevitably occur. A confidence of 95% signifies that one in twenty innocent defendants will be wrongly convicted. Any identification of a threshold would explicitly acknowledge that the system officially condones a certain fraction of wrongful convictions, which would supposedly threaten the ordinary person’s faith in the criminal justice system.’* Friedman [26] explains that society’s unwillingness to accept explicit flaws stems from fundamental principles of ‘utility’: *‘Blackstone’s [c.1760] injunction that it is better to let 10 guilty persons go free than to convict one innocent person may well be a substantial understatement. This utility*

assessment leads to the 'beyond a reasonable doubt' standard in criminal cases.' Utility is based on pragmatism, or in other words 'making the best of the situation'.

Laudan explains that there is implicitly a balancing act performed to ensure that flaws in the system are kept to a bare minimum: *'Every increase in the standard of proof makes it harder to convict the truly guilty. Unless the overwhelming majority of those who come to trial are genuinely innocent, raising the standard will make false convictions more likely than they would otherwise be. A very high standard is bad news in terms of a crime deterrent or retribution for wronged parties. Once we have settled on a socially acceptable ratio of true acquittals to false convictions, we know what we are looking for in a standard of proof. A standard set at 91% would ensure ten true acquittals for every false conviction....If the standard is set at 90%, the question to a juror would be: 'is your assessment of the likelihood of the Defendant's guilt higher than 90%? If so you must convict.'* Instead of specifying that the juror's level of confidence in guilt should depend on whether a robust proof has been offered, the criminal law makes the standard parasitic on the juror's level of confidence in the Defendant's guilt. We have proof, so long as jurors are strongly persuaded of the guilt of the accused. This gets things precisely backwards. This would be like saying to a mathematician that they have proof of a theorem if they are convinced the theorem is true; or to an epidemiologist that they have proof of a causal link between A and B if they are convinced such a link exists.' This balancing exercise between acquitting a guilty person and convicting an innocent one was addressed by Tribe in 1971 in criticising the Kaplan/Cullison 'Jury Decision Model' as a possible formulaic alternative to the current 'jury guessing' method of deciding guilt.

Friedman explains that a utility model, such as that proposed by Kaplan and Cullison, would help improve transparency in the decision-making process: *'Suppose for simplicity that we are limited to two options - finding for the plaintiff or finding for the defence. Then, at least as a first approximation, it seems that a wise guide to decision is to choose the option that has the greater expected utility. Expected utility will depend on the relative probabilities that the facts support the plaintiff and the defence and on the utilities attached to each possible decision given each possible actual state of affairs.'* In this way, when a value lower than 1 is provided by the model, it will reflect the true decision-making process of the fact finder. It is this aid to transparency which makes the utility model so attractive, because no longer can it be said that uncertainty is unquantifiable.

The only difference between a utility model and a jury-guess decision is that this uncertainty is elucidated. This surely must be in line with the legal community's desire to encourage transparency in every stage of the justice process. The only issue now is 'where below '1' should the line be drawn'?

Edwards [12] asserts: *'I suggest that posterior odds of 100:1 might be an appropriate, though demanding, operationalisation of 'beyond a reasonable doubt... The adoption of, say, 100:1 posterior odds cut-off would not result in one in a hundred convicts actually being innocent. Prosecuted cases will rarely fall at the cut-off; most will be beyond it. Cases that, in the prosecution's prior assessment, will not reach the cut-off presumably would not be prosecuted. We cannot know exactly how many innocent convicts such a definition would lead to, but we can be sure that it would be very much less than 1 in a 100. A change in the cut-off from 100:1 to 1000:1 would make a difference only in cases for which the posterior odds fall in that range – presumably a very small subset of the total set of cases.'* Of course, the problem with this approach is that the 'very small subset of cases' suggested by Edwards is arbitrary and still will not sit well with those who consider a repugnant 'trial by mathematics' a natural progression from setting objective posterior cut-offs.

However, not every member of the legal community is resolutely against broaching the issue of an objective posterior limit. Tillers and Gottfried [31] explain that in the U.S., Judge Weinstein has advocated the use of an objective posterior to reduce arbitrariness in legal decision-making: *'As Weinstein suggested in Copeland III, one possible powerful argument for some sort of numerical quantification is the importance of the uniformity of legal standards. Weinstein correctly suggested that differential legal standards are particularly disturbing when they appear in the criminal justice system. If numerical quantification of the burden of persuasion in criminal trials can reduce differences in the operational meaning of the standard of persuasion in criminal trials, a powerful argument in favour of numerical quantification is at hand.'* Franklin [42], in also discussing Judge Weinstein's views, does not advocate a positive number, but argues that any decision falling BELOW a particular threshold should not return a verdict of guilt: *'An appropriate numerical standard to choose as an absolute minimum follows from Judge Weinstein's suggestion in 'Copeland III' (United States v. Copeland, 369 F. Supp. 2d (E.D.N.Y. 2005) of 20% for a 'reasonable probability', and hence of 80% for its inverse or complement 'clear, unequivocal and convincing'*

evidence. Since proof beyond reasonable doubt is well above clear, unequivocal and convincing evidence, it follows that proof beyond reasonable doubt means 'well above a probability of 0.8'. Any suggestion from a jury that 0.8 or less is adequate can be ruled out, while the qualification 'well above' will avoid any suggestions that something just above 0.8 is in fact adequate, and will not obstruct any later attempts to quantify the standard more exactly.' What is interesting here is that Franklin acknowledges that the debate over posteriors must start somewhere. By suggesting a lower 'safe' threshold, the debate can move beyond a polarised 'yes'/'no' argument on introducing an objective cut-off, to a more nuanced debate about exactly where in the scale the cut-off should lie.

What cannot be denied is that juries often struggle with the concept of BARD, if left alone to value it. Judges in the UK sometimes refer to the standard as a 'moral certainty', which is arguably no more helpful than the current standard. Tillers and Gottfried [31] argue that attempts are routinely made to explain the criminal standard of proof and provide some sort of framework for establishing the value of reasonable doubt: *'Prosecutors and judges often use jigsaw puzzle analogies to explain the assessment of circumstantial evidence, which the appellate courts do not usually frown upon, as long as the analogy does not refer to specific numbers of missing pieces. In one of those cases, the trial judges instructed the jury to compare circumstantial proof beyond a reasonable doubt to a '1000 piece puzzle with sixty pieces missing'.* This issue of fact finders being given an objective cut-off, rather than they themselves formulating their own, seems to be at the very centre of the argument against a utility model.

The answer is to accept that the cut-off is subjective to each and every fact finder, and to allow them the opportunity to use the model to simply aid them in deciding whether their own posterior arrived at correlates with their views on the guilt of the defendant. Laudan explains that this subjectivity of BARD cannot be denied: *'A study by Cohen and Christensen in 1970 interviewed English judges and jurors about the level of probability that should be required for conviction in a criminal trial. Among judges, a third located it between 0.7 and 0.9, with almost all of the rest putting it above 0.9. Among jurors, 26% were willing to convict on probabilities lower than 0.7, and a 54% though that above 0.9 should be required for conviction. The idea that the standard of proof is set at such different points on the scale by judges and jurors should be a grave worry.'* Redmayne [18] agrees: *'The approach of judges in the United Kingdom to instructions on the meaning of beyond reasonable doubt*

now appears to be to avoid attempts to “define that which is almost impossible to define”. Academics, though, have not felt so daunted by the criminal standard. Indeed, at least one attempt has been made to obtain a numerical expression of beyond reasonable doubt. However, such attempts ignore the fact that reasonable doubt is an inscrutable, subjective standard which exists largely within the mind of the fact-finder in a particular case.’

The Crown Court Bench Book ‘*Directing the Jury*’ [37] advises judges that the standard of proof in criminal cases is as follows: ‘*Standard of Proof: The prosecution proves its case if the jury, having considered all the evidence relevant to the charge they are considering, are sure that the defendant is guilty... Note: Being sure is the same as entertaining no reasonable doubt*’. Since ‘sure’ means ‘no reasonable doubt’, is this a variation on ‘beyond all reasonable doubt’, and, if so, what is the difference?

Archbold Criminal Pleading Evidence and Practice [43] states that there is no difference between the ‘sure’ test and BARD test; at paragraph 4-444: ‘*...before the jury can convict they must be satisfied beyond a reasonable doubt (or be sure) of the defendant’s guilt*’; but in *R v Summers* [44], Lord Goddard C.J. merely stated that it was better to tell the jury that before they convict they must be ‘*satisfied so that they are sure*’ of the guilt of the accused, on the basis that some judges had found ‘*difficulty in explaining what was meant by “reasonable doubt”*’.

In *R v Bracewell* [45], the trial judge summed-up the expert witness, Dr Green’s, evidence thus: ‘*You will remember ladies and gentlemen that your duty is not to judge scientifically or with scientific certainty. You judge so that as sensible people you feel sure and even say that what might not satisfy Dr. Green as a scientific certainty, might with propriety, satisfy you so that you felt sure. Do not be misled. There is no such thing as certainty in this life, absolute certainty. You ask yourselves the simple question upon the whole of the evidence do I feel sure?*’ On appeal, the Court of Appeal held that the direction was correct: ‘*That direction, in our judgment, correctly draws the distinction between what might be described as scientific proof on the one hand and legal proof on the other*’. This is a vital distinction – not only did the trial judge direct that the ‘sure’ test was subjective (‘*...so you felt sure*’), but the Court of Appeal endorsed the difference between expert evidence (‘*scientific proof*’) and the jury’s degree of belief (‘*legal proof*’) in that evidence.

It can therefore be seen that whether the standard of proof is ‘no reasonable doubt’ or ‘beyond all reasonable doubt’, the jury must be left to decide whether the evidence is enough to convict. The criminal standard of proof test is therefore ultimately subjective to the jury.

To sum up; again it is obvious that Tribe’s argument against a ‘trial by mathematics’ – in this case due to BARD being irreconcilable with quantified probabilistic reasoning – cannot be sustained. The standard of proof is subjective and down to each fact finder individually to establish, using their own pragmatic concept of ‘utility’, if the threshold has been reached. While objective quantification of the standard is undesirable on policy grounds, it cannot be said to stop fact finders using the posterior as a means to start the decision-making process.

If on the evidence in a case a juror discovers, after using a model, that the posterior is 95% and wishes to return a verdict of ‘guilt’, it does not mean that five in every hundred convicted people in prison are innocent. It simply means that on the facts of *that particular case* the juror decided that guilt was the appropriate verdict. In other cases, a juror facing a posterior of 95% may decide on utility grounds that the defendant should be acquitted. Either way, it cannot be argued that the use of a model would not be a useful tool in evaluating the strength of prosecution cases, and would provide useful information about the decision-making processes and reasoning of the fact finder in criminal trials.

Essentially, the problem with the ‘Blue Bus’ case, is not that the bus company would be unfairly found guilty even though there was a weak/no causal link between itself and the alleged event, but because the legal notion of the standard of proof is wrong. Tribe’s stance on this point actually creates MORE uncertainty than it resolves.

On a final note, in a jury of twelve members, the minimum majority required for a successful verdict is 10-2. This equates to a maximum of 83% certainty, even if the ten who return a verdict of guilt are each 100% certain of guilt (an outcome which is vigorously refuted here, as 100% certainty is impossible to achieve). While it is controversial to accept that 83% should be considered as a viable threshold across the whole range of the jury, the point made is that even under the current system, a degree of uncertainty is built in.

5.3.5 (v) That due to the complexity of cases and non-sequential nature of evidence presentation, any application of Bayes' theorem would be too cumbersome for a jury to use effectively and efficiently:

In technological terms, Tribe's 40 year old argument - that a jury's use of Bayes' theorem (BT) would be 'too complex' for ease of application - MUST be outdated. Even if the judgment in *R v Adams* [46] a quarter-century later in 1996, which held that: '*...[W]e do not consider that [juries] will be assisted in their task by reference to a very complex approach which they are unlikely to understand fully and even more unlikely to apply accurately, which we judge to be likely to confuse them and distract them from their consideration of the real questions on which they should seek to reach a unanimous conclusion*', is still binding on lower courts today, it cannot be true that modern computers cannot eradicate much of the cumbersome, in order that applying BT to even complex cases becomes easy and commonplace.

Advances in technology, used to lighten much of the mathematical load. was recognised by Edwards [12] as long ago as 1991: '*The proposal that trial lawyers should exploit Bayesian tools makes a lot more sense now than before. Bayesians are aware that conditional independence is rare, that in its absence, judgements of likelihood ratios must be replaced by more difficult judgements of conditional probabilities and that Bayesian tools used to be too cumbersome for routine use. That has changed, thanks to influence diagrams and Bayes nets. The assessment task, of conditional probabilities, can be given an orderly structure, the task of specifying relevant Bayesian equations can be made invisible to the user, and Bayesian arithmetic can be automated by the use of INDIA and HUGIN computer programmes.*' Of course, the issue in *Adams*, five years after Edwards' recommendations, was not of the jury's use of the mathematical tool, but the confusing (if not error-filled) way in which the witness presented the evidence.

These days, much of the presentation would be handled through Bayes' nets (BNs). Bring's 1997 argument that [29]: '*I think the eminent work by Kadane and Schum on the Sacco and Vanzetti case proves that Bayes' theorem can be an interesting method of studying evidence but that it is impossible for normal fact-finders to use these methods in practice*' must also be out of date. Friedman [26] acknowledges that BNs are more than up to the task of handling the relationships between complex evidence and multiple hypotheses: '*Some Bayesioskeptics*

emphasise the computational complexity created by the Bayesian approach. I think the argument is wide of the mark, for several reasons:

(i) The world is indeed a complex place, but that does not reflect a problem with Bayesian analysis. On the contrary, any theory that could not in principle represent the complexity surrounding us would have limited value;

(ii) I say 'in principle' because the theory need not be applied in its most powerful gear. On the contrary, it is a flexible template. It can take into account as much complexity as its user is able to handle;

(iii) I think the computational complexity, though great, may not be as great as it appears. The storytelling Bayesian approach that I have outlined above is far simpler than an item-by-item approach. It requires a limited number of probability assessments, and does not demand that the observer cross-check for consistency all the probabilities that she might assess given multiple items of evidence;

(iv) One model requires simply that an observer adhere to the following rule: If $P(E|H) > P(E|NH)$, then $O(H|E) > O(H)$, and the greater the proportional difference between the first pair the greater the proportional difference between the second pair. This rule follows immediately from Bayes' theorem.

One could comply with this rule without complying with the theorem, but usually the rule would tend to lead to results rather close to results prescribed by the theorem. While Friedman advocates that BT is not too complex a theory for use by jurors: 'I think the Adams court was off base in suggesting that Bayes' theorem has no place in the courtroom....' he does argue that the 'inner workings' of the theorem are not important to be divulged: 'On the other hand, as I have already indicated, I tend to think that it is not helpful to take a jury through an iterative use of Bayes' theorem, even in a case like Adams in which the evidence was very sparse.' This is the cornerstone of the argument for the modern implementation of BT – a stance which has led Fenton and Neil [47] to support the 'Electronic Calculator Theory': 'However, we disagree with the Royal Statistical Society's stance that: '...statistical evidence is presented only by appropriately qualified statistical experts, as would be the case for any other form of expert evidence', nor do we suggest that lawyers and juries should be trained to do the calculations themselves, as suggested by Robertson and Vignaux in 'Don't teach Statistics to Lawyers!', in Proceedings of the Fifth International Conference on Teaching of Statistics' and 'Explaining Evidence Logically'. Instead, we suggest that BNs should be used in the same way as an electronic calculator is used to provide everyday

calculations.’ With this theory, BNs, and the software that applies them, are used much in the same way that an electronic calculator has become a commonplace tool, used by non-mathematicians.

There is much to be said for this approach – certainly, it can be said that laypersons do not question the circuitry or foundation principles of the calculating capabilities of the modern calculator, so why should there be an innate distrust of BN software? Fenton and Neil explain how a jury would be directed to use the BN software [3]: *‘After a BN model was used to calculate the value of the evidence, the lawyer in a criminal trial might say: ‘What we have demonstrated to you is how we revise our prior assumption when we observe a single piece of evidence. Although we were able to explain this to you from scratch, there is a standard calculation engine (accepted and validated by the mathematical and statistical community) which will do this calculation for us without having to go through all the details. In fact, when there is more than a single piece of evidence to consider it is too time-consuming and complex to do the calculations by hand, but the calculation engine will do it instantly for us. This is much like relying on a calculator to do long division for us. You do not have to worry about the accuracy of the calculations; these are guaranteed. All you have to worry about is whether our original assumptions are reasonable. But we can show you the results with a range of different assumptions.’ The lawyer could then present the results from the BN model and explain that the results exactly match the results in an event tree. First, the lawyer shows the results with no evidence entered, then the DNA, then, the eyewitness evidence, then the alibi, showing how the values change with each piece of evidence. A range of values could be used to give upper and lower parameters within which the jury would work. It would then be for the jury to decide whether the assumptions in the model are reasonable.’* This argument supports Edwards [12] views from 1991 who suggested that in a ‘four-tier’ BN node, the lowest tier would be inaccessible to the user and would contain the necessary algorithmic software which would act as the ‘engine’ of the BN. This discussion has been more fully introduced in Chapter 2.

Of course, the problem of complex application of a simple theorem is not the only problem in play here. The other problem of ‘complexity’ concerns the accuracy of values placed INTO the model. Can a juror be trusted to clearly elucidate his own reasoning simply and effectively enough to be able to place accurate figures into the calculator.

Taking the Electronic Calculator Theory to its natural conclusion, where a juror may know that adding three items to two items can be transferred to a calculator's key pad efficiently as $3+2=5$, can a juror just as effectively transfer knowledge of a complex case with multiple pieces of evidence to a similar keypad. Would he know what values to key in? Ronald Allen [48] says no: *'The point is that the problems humans typically face at trial are inferentially complex, too complex to permit the use of any version of Bayes' theorem as either an algorithm or a heuristic.'* Allen then argues that not only can complex cases be filled with many different pieces of evidence, there may also be contradictory hypotheses which may cause confusion in a juror not instructed on how to use a calculator: *'The conventional legal argument assumes that there are only two relevant hypotheses: guilt and not guilt. In many instances, there are multiple competing hypotheses. In that case, the complexities increase significantly.'* In any case, Allen goes on to say that even after the case has been fully presented by both prosecution and defence teams, the calculator would need to be constantly updating the reasoning process: *'During deliberations, in some cases, maybe in many, new theories will emerge. When they do, the probability space must be reconfigured, and there is no algorithmic way of doing so.'* Would this constant use of an electronic device deflect the juror from his proper task, as was warned about in *R v Adams (Denis)* [46]: In that case the Court of Appeal held that: *'It seems to us that the difficulties which arise in the present case stem from the fact that, at trial, the defence were permitted to lead before the jury evidence of the Bayes Theorem,'* and later stated that: *'...the apparently objective numerical figures used in the theorem may conceal the element of judgment on which it entirely depends... it seems to us that it is not appropriate for use in jury trials, or as a means to assist the jury in their task... Jurors evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them... as the present case graphically demonstrates, to introduce Bayes Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task.'* In that case, a rape trial where the identity of the suspect was at issue and, much like in the US case of *Collins*, the prosecution case included probabilistic reasoning to illustrate the probability that the defendant was the suspect, the prosecution expert, Professor Donnelly, explained in detail the figures he used in the mathematical model and the methodology used to provide the likelihood ratios.

The trial had ruled that BT was available for use by the jury and stated: *'If you do not wish to use it that is your privilege and your own private decision and no one will criticise you for not using it. There is absolutely no compulsion on you to use it at all. It is there if you want to use it and follow the instructions given.'* However, the trial judge, in summing-up, misdirected the jury when he became confused by the expert evidence and, in particular, appeared to have forgotten the answer given by Professor Donnelly expressing the probabilities as percentages. Instead, the judge was directing the jury as to how many times it was more likely that something had occurred, yet he did not remind the jury of the formula given by Professor Donnelly in relation to the percentages.

However, let us say that Professor Donnelly and the trial judge had applied BT in a simple and formulaic way and then left the jury to deliberate. Can it be said that the jury, in the privacy of the deliberation room, could be trusted to apply BT correctly even with modern software and the correct judicial directions proposed by Fenton and Neil, and supported by Jowett [49] *'It should be acceptable simply to provide the results of Bayesian arguments arrived at using tools such as Hugin. Experts would then testify that the tool has been used correctly.'* Allen's argument that the model would need constant updating may mean that an expert would need to be on hand in the deliberation room to assist – which could mean that the essential privacy and potential autonomy of jury deliberations would be compromised.

To sum up; if the assertion of Allen [28] , that computer complexity is the greatest threat to the implementation of BT in court, is correct: *'The primary problem is that Bayes' theorem cannot be implemented in a typical trial, for a number of reasons such as computational complexity'*, and it is accepted that with fairly recent great strides having been – and still being – made in Bayesian software programmes, then Allen can no longer sustain one of his main objections to BT. In fact, nowhere in Allen's published works is any consideration given, serious or otherwise, to the workings or operation of a BN.

However, what cannot be ignored is that in some highly complex cases, it must be accepted that once a jury is left to deliberate, that it cannot be acceptable to have a third party assist with re-configuring the software to encompass scenarios not autonomously considered by the jury itself. Therefore, while fairly simple cases which do not need outside interference by an expert into the deliberating jury's domain may be suitable for an application of BT, the line must be drawn at cases which compromise the autonomy and privacy of the jury's role. Of

course, this does not mean that the software should not be made available to the jury for its own use, but that the instruction period must end before deliberation.

5.3.6 (vi) That probabilistic reasoning is not compatible with the law, for policy reasons:

The argument that probabilistic reasoning is not compatible with law for policy reasons is based on the idea that it would be somehow repugnant to society to reconcile the idea of quantitative mathematical proof with the age-old system of unaided jury deliberations.

Allen [48] argues that since legal trials encompass more than simply a quest for truth about a given hypothesis, that a probabilistic model can do nothing to create hypotheses: *‘If the law is not Bayesian, what is it? The probability debates have proceeded as if the law deals only with uncertainty; in fact, it deals as much, maybe more, with ignorance. The typical litigated case is not like assessing the probability of obtaining a certain number of heads of an evenly weighted coin, given a certain number of flips of that coin. It is not, in short, a case of uncertainty (given the definition above). It requires instead determining which one of a very large range of possibilities actually occurred at the time in question in the face of missing, disputed and contradictory evidence. The typical litigated case is like the urn of ignorance.’* In short, this argument is closely aligned to the ‘soft variables’ argument discussed above, which discusses the idea that if a probability cannot be counted, it does not exist.

However, what Allen is really arguing here is that a mathematical model can never truly perform the task it is purported to be designed for – namely that of proving guilt. This task must be performed by a jury, which although may be a flawed model, is the best model available. Allen explains the impossibility of reconciling BT with the jury’s task: *‘Reconsider the standard use of Bayes’ theorem. It conditions guilt on the evidence, but guilt is a legal conclusion. It is not a fact, nor a theory about facts that can be reduced to probabilities. It is a description of an act taken by a legal actor.’* The problem with this approach is that Allen considers that ‘guilt’ is unquantifiable, much like a ‘feeling’ or an ‘emotion’.

These human qualities, Allen argues, are a necessary part of the criminal justice system. Of course, since feelings and emotions are part of the flawed human condition, it would be

wrong to pin a value on them for fear of losing something in the translation. Tillers [50] does not adhere to this view: *‘The generally unproductive and sterile debate about whether there should be a trial by mathematics was founded on two misunderstandings: (i) A widespread failure to appreciate that mathematics is part of a broader family of rigorous methods of oft-called ‘formal’ reasoning; and (ii) A widespread failure to appreciate that those mathematical and formal analyses can have a large variety of purposes.’* What Tillers argues here is that mathematical reasoning is simply a branch of any type of human reasoning, which must include jury deliberations.

Friedman [26] supports Tillers view that probabilistic reasoning is simply a quantitative version of what is already present in the mind of the juror, which means that banning it would be futile: *‘Recent research suggests that fact-finders tend to view an entire body of evidence, attempting to determine a story that most plausibly accounts for all of it. Some Bayesioskeptics have viewed this model as conflicting with a Bayesian view of evidence. I do not believe there is any genuine conflict. I do not believe that fact-finders can, should or do go through such a serial updating of probability, given each new piece of evidence. How, then, can a Bayesian approach operate on a whole body of evidence? Suppose the fact-finder, in choosing between $H1$ and $H2$, has four pieces of evidence, $E1$ having arisen beforehand and $E2$, $E3$, and $E4$ having arisen afterwards. Then the fact-finder may compare competing stories, one for $H1$ and one for $H2$ that account for all of the evidence. This is a story-telling model, but it is also Bayesian.’* This approach must be correct. The difference between Friedman’s argument and Allen’s therefore, is that Friedman is willing to delve into Allen’s ‘urn of ignorance’ and tease out the hypotheses from within. This strive for a formulaic approach to legal reasoning conforms to the legal community’s strive for transparency and acknowledges flaws and tries to do something about them, rather than simply accepting they exist and ignoring them.

Allen tries a different approach: *‘In science, the problem of priors is supposedly satisfied by what are called convergence-to-certainty and merger-of-opinion results. Over time, as evidence is produced, opinions will converge on the truth no matter where the initial starting points happened to be. But, this bears no relationship to jury decision making. Fact-finders do not run further tests to sort out the theories. They have to make decisions based on what they have. This point is a dramatic demonstration of the apparent failure of the Bayesian enthusiasts to attend to the foundations of their own theory. It is just these theorems involving*

convergence-to-certainty that allow a plausible case to be made about the use of Bayes' theorem in scientific inference, because they connect the subjective to the objective. Over time the stability of nature will wash out the subjective starting place of humans. What is the analogous connection between the subjective and the objective in the law?' This argument follows Tribe's original claim that to allow probabilistic reasoning into court would lead to a 'trial by mathematics', thereby leading to a usurping of the jury function.

In Allen's argument, the 'subjectivity' in a model - vital to preserve the jury's role - would be lost, and the machine's objectivity (no doubt programmed by a small group of experts, now in charge of all trial deliberations) would be left to prove guilt. Of course, this scenario may be far-fetched, but it is a problem which has already been addressed. The human fact finder's role is a vital one in the criminal justice process. Over the past 40+ years since Tribe's article, computers have pervaded every aspect of human existence, from assisting medical procedures to going to war. The argument that using computer technology to assist in decision-making tasks would somehow usurp the human element has been found to lack foundation. As long as BT is merely used as an aid in the juror's decision-making process, there can never be a 'trial by mathematics'.

5.4 Conclusion:

As has been shown here, there is very little in Tribe's 1971 paper which stands up to either (a) scrutiny; or (b) the test of time. While the science of trace evidence matching has evolved towards the commonplace introduction of the RMP to the trial process, the UK courts reluctance to embrace explicit quantitative probabilistic methods to meet those demands must be examined, due to its irrationality.

In doing so, the parallels between the Nuffield Foundation's anti-Bayesian stance towards most types of evidence and the UK Court of Appeal rulings, are marked. In effect, the contents of Laurence Tribe's 1971 paper remain almost untainted by those rulings, despite massive advances in the deeper understanding of the basic principles of probability theory – taking it within the knowledge of more than just a handful of practitioners – and advances in computer science - to handle the complexity of Bayesian modelling – which should logically have made the introduction of BT into the criminal trial process as prolific as the RMP itself.

However, for reasons explained and alluded to in this chapter, this has not happened, and the UK courts for one reason or another have decided that BT remains a niche area of discipline, amenable to only the narrowest of circumstances.

6 CHAPTER VI: Priors

6.1 Introduction:

The focus of this chapter examines how to formulate the correct prior for any given hypothesis. This is not an easy task, yet there has never been an in-depth academic study into the mechanics of the ‘legal’ prior - which is surprising, given the central role it plays in any probabilistic model.

The first part of this chapter explains the background to the problem, and broaches the subject of whose task it is to assess the prior. Since a forensic study of the inner-workings of the prior are necessary to pinpoint the problems and controversies inherent in assessing the prior, a full definition of the prior, and how one is formulated, is given.

The chapter continues by addressing the value ‘ K ’ (which is mathematical notation for ‘context’), in order to explain that any mathematical model must consider the context of the hypothesis and its connecting evidence to ensure that the model provides accurate information relevant to it. Again, this is an area which is barely discussed among the mathematical community - only touched upon by Fenton & Neil [3], and never discussed by the legal community - which is also surprising given that it is a vital component in any probabilistic reasoning process.

After the discussion and explanation of ‘ K ’, there is a discussion about trace evidence – of which DNA falls within this type – and an examination of the random match probability (RMP) which is now routinely used as a tool for presenting this type of evidence to the jury. This discussion leads to the unveiling of the revolutionary ‘1/World Population’ (1/WP) prior for all types of trace evidence – a prior which has never been considered before, and which should counter all arguments against the courtroom use of Bayes’ theorem for cases where the identity of the suspect is the key issue.

The chapter then provides illustrations of how evidence can be combined in a single model, by breaking down decision-making stages into composite parts in order to expose misconceptions and errors in the way that juries might combine evidence which overlooks the

importance of 'K' - thereby allowing evidence to be ascribed inaccurate weight, which will naturally increase the risks of miscarriages of justice.

Finally, the chapter provides basic examples of Bayesian networks (BNs) which may offer a solution to the problem, and provides crucial support for the research hypothesis.

6.2 Assessment of the prior:

If the probabilistic reasoning community are successful in their bid to implement probabilistic modelling in criminal trials, it must prove to the legal community that the model in no way interferes with the jury's role of ascribing weight to evidence and returning a verdict. [51] *'The important point is that assessment of the prior odds is the jury's task'*. [18] The model works simply on a principle of combining prior probabilities with evidence, to calculate a posterior probability. And therein lies the first problem. How does a juror formulate a prior without observing any evidence?

Johan Bring [52] identified the issue of the difficulty in arriving at a rational single point prior, and Fenton & Neil [3] sought to resolve the problem by suggesting that a juror choose a 'range prior' encompassing and most closely fitting his own view of the weight of a single piece of evidence. This range prior was originally suggested by Finkelstein & Fairley [23] as a way to avoid a third party assigning single point values to a juror, thereby avoiding any accusation of interference.

Ronald Allen [28] justified the use of subjective priors in non-legal sciences, such as medical testing, due to the scientifically collaborative nature of these disciplines, but suggested that in criminal law, the same collaboration is not in attendance due to the juror's lack of knowledge of the case, that therefore the '*convergence-to-certainty*', or '*merger-of-results*' principles which are prevalent in medical testing and act as the necessary bridge between the subjective and objective, are not present in criminal trials. What this means is that regardless of how much subjectivity is superficially afforded the juror, ultimately the model in criminal trials is objective due to the juror's lack of opportunity to autonomously formulate his own prior from data which he has some degree of control over. James Berger [53] advocates the use of objective Bayesian analysis in narrow appropriate situations, but the legal community will, quite rightly, not allow any mathematical model to interfere with the jury's autonomous role.

The legal community's reluctance to accept probabilistic reasoning in criminal trials can be traced at least as far back as 1968 and the Los Angeles robbery case of *People v Collins*. [5] Laurence Tribe [4], in his post-case critique of the prosecutor's mathematical reasoning (where it is inarguable that the prosecutor committed the prosecutor's fallacy [15] among other fundamental mistakes), argued that the prior was a potentially controversial area due to the jury's difficulty in arriving at (a) an accurate figure; and (b) a number which would not overpower other types of '*more impressionable*' evidence.

In the 1996 case of *R v Adams (Denis)* [54], the expert witness for the prosecution suggested that the prior should be formulated from a geographical sub-population, or 'reference class', which when combined with the random match probability (RMP) of a piece of DNA evidence found at the crime scene, would provide an overall posterior probability of guilt. Due to the overly complex method of presenting the probabilistic reasoning, the court ruled that Bayes' theorem (BT) would '*deflect the jury from its proper task*' and that BT should not be recommended for use to a jury as a matter of course. However, this formulation of a prior from geographical sub-populations, or 'location evidence', has been accepted by the scientific community, including Fenton & Neil who recently confirmed and recommended combining evidence of sub-populations with DNA RMPs to provide inferences of guilt. This is echoed by Buckleton et al [6] who believe that DNA is a valuable tool in establishing guilt, but only when used in conjunction with other types of evidence, including location evidence.

The Crown Prosecution Service (CPS), which brings cases based on strength of prosecution evidence, also recommend that DNA should be used in conjunction with location evidence, such as '*a starting point might be the British population*', or '*the sexually active male population of Manchester*', and recommends that '*...to do this, the scientist would 'need to be given a guide with regard to the size and nature of the population to consider*' [7]. It suggests that in a rape case in Huddersfield '*an old woman from Beijing*' sharing DNA profile traits with that of the crime scene trace would likely be summarily excluded from enquiries, on the basis that the probability of her being the suspect was so low as to be rationally disregarded. This indicates that the CPS value DNA evidence, but would likely only bring a case on an amalgam of different types of evidence - with DNA only being a single type and therefore incapable of sustaining a guilty verdict on its own.

The question is *can* DNA be combined with other types of evidence? This question is directly related to the enquiry of what the appropriate priors are for different types of evidence, and whether these priors can be combined to calculate a single holistic posterior value of guilt. Redmayne [18] argues that values of ‘other types’ of evidence should provide the prior for a combination model, before DNA is introduced: ‘*The prior odds should be assessed on the basis of any other evidence against the suspect before the DNA evidence is introduced.*’ However, is this actually possible?

6.3 What is a ‘prior’?

BT is a method for updating our beliefs about a given hypothesis. By combining evidence with our prior belief, we become more certain in the hypothesis until we feel ready to make a decision related to that particular hypothesis. If my hypothesis is that ‘the weather requires me to carry an umbrella’, the evidence of rain clouds in the sky may lead me to a decision to take an umbrella with me when I go to work. Our prior beliefs are usually formulated from our view of the world as a whole, which is, in turn, created by our own subjective experiences. In this way, BT is very personal to the fact finder who is using it as an aid to his decision-making.

Fenton & Neil argue that this subjectivity is present in all probabilistic reasoning, since ultimately any information said to be ‘objectively accurate’ is only as accurate as the data collection and results interpretation methods implemented. Ward Edwards [12] even argues that a person’s height is subject to individual interpretation due to the subjective reliance placed by the fact finder on the integrity of the measuring tape. This ‘leap of faith’ is the foundation of the subjectivist view on probability and has not been adequately rebutted [47] by frequentists, such as members of the Court of Appeal in cases such as *R v T* [2], who argue that it is theoretically possible to have an ‘*objectively certain*’ database.

Since BT is used to reduce uncertainty in a given hypothesis, it has been suggested [3] that it is a good fit with criminal trials. This is theoretically true. BT is a simple formula which, if used appropriately, can increase transparency in the decision-making process and reduce miscarriages of justice by eradicating probabilistic reasoning errors and fallacies. The problems, however can arise where the prior probabilities have not been correctly formulated

and communicated. This element of communication, normally irrelevant in non-legal disciplines, is vital in criminal trials, because the trial fact finder – the juror – and the evidence collector – the witness - are not the same person.

Therefore, the probative value (the definition of ‘probative value’ can be found in Chapter 2) of a piece of evidence can be misconstrued or misinterpreted if the witness does not adequately explain to the juror what the evidence is and what it means to the case as a whole. For example, a bloody knife in a murder trial could either be evidence of the defendant’s ‘inculpability’ as a murderer, or of his ‘exculpability’ as a murderer - if it has been used as evidence of his job as a non-murderous butcher. Simply laying the knife on the exhibit table and asking the juror to make up his mind as to the meaning of the knife’s presence in court is fraught with the risk of misinterpretation. The same can be said about DNA evidence. In the conjoined appeals of *R v Adams & Doheny* [1] the ‘random match probability’ (RMP) was introduced as means of presenting the probative value of DNA found in a crime scene trace.

The concept, is that the ‘probability of seeing the evidence if the prosecution hypothesis of ‘guilt’ $P(E|H_p)$ is correct’, is then divided by the ‘probability of seeing the evidence if the defence hypothesis of ‘innocence’ $P(E|H_d)$ is correct’, to arrive at the probative value of the evidence. Since, in Doheny, the DNA was stated by the expert as being prevalent in 1/27 million random people in the world population, the actual number of matches would be around 250 people, giving a posterior probability of guilt of $\approx 1/250$. However, the prosecution expert witness transposed the conditional – confusing $P(E|H_d)$ with $P(H_d|E)$ – thereby wildly overstating the prosecution case - which the jury believed and agreed to be a posterior probability of *innocence* at an extremely low 1/27 million, and convicted the two defendants on the basis of the prosecutor’s fallacy (as discussed fully in Chapter 3) that:

$P(H_d|E)$ = The probability of the defendant being innocent, given the evidence, is 1/27 million (a very low probability of innocence)

...which was confused with:

$P(E|H_d)$ = The probability of seeing the evidence, given that the defendant is innocent, is 1/27 million (a high probability of innocence)

This misinterpretation of the RMP was simply made on the basis that the prior probability of guilt was not properly communicated between witness and jury. If it had, then the BT calculation of: $P(E|H_p) / P(E|H_d) \times P(H_p) / P(H_d) = P(H_p|E) / P(H_d|E)$ would have been unnecessary. This miscommunication relates to the hidden value ' K ' in this, and any, probabilistic model.

6.4 The value ' K ':

Fenton and Neil's argument that all probability is subjective is underpinned by a principle that *'probability assigned to an uncertain event A is always conditional on a context K , which you can think of as a set of knowledge and assumptions.'* This value of K is inherent in all probabilistic values and is present in BT, but often overlooked or taken for granted. The 'odds' form of BT, with K expressly stated, would be:

$$P(H_p|K, E|K) / P(H_d|K, E|K) = P(E|K, H_p|K) / P(E|K, H_d|K) \times P(H_p|K) / P(H_d|K) .$$

Of course, because of its cumbersome nature, K is cancelled out across the formula and taken to exist 'behind the scenes'. Problematically, K , once removed from the formula, is often forgotten.

In criminal trials, this silent value K is highly valued due to the symbiotic relationship between the two fact finders, witness and juror; but in non-legal disciplines, the fact finder is usually a single person, who is innately aware of K and instantly factors-in its value wherever necessary. If K was expressly communicated by the witness to the jury at the outset of criminal trials, the prosecutor's fallacy would not exist, although it is easy to see how mistakes can happen: Taking the prosecutor's fallacy as an example of how a RMP might be interpreted:

Hd: The defendant is not the source of E

E: DNA with a RMP of 1/1 million

$$P(E|H_d) = 1/1 \text{ million}$$

This can be interpreted in two different ways:

1. 'The probability of seeing the DNA evidence if the defendant is not the source is one in a million'
2. 'The probability of seeing the evidence if the defendant is not the source is one in *every million people*'

The addition of the words '*every million people*' in the second example make the difference between a strong prosecution case and a weak one, yet unless these words are expressly communicated by the witness to the jury, it is easy to see how a juror might confuse the examples and attribute an inaccurate value to the evidence. In essence, an abstract value of 1/1 million if presented without value K , is highly prone to misinterpretation.

The Nuffield Foundation [16] recognises the importance in communicating value K to the jury when presenting evidence: '*When one grasps that evidence (including expert evidence) is adduced by the prosecution or defence to answer a particular question, it follows that the meaning and value of that evidence cannot be determined without first identifying the original question. One cannot assess whether evidence is successful in proving a matter in issue until one knows what the issue is and how the evidence relates to it. This observation might sound banal; but it is not. In fact, nearly all of the reasoning errors are either variations on, or are at least exacerbated by, an elementary failure to identify, with sufficient care and particularity, the question which the evidence is capable of answering.*' This nexus between 'the question' and the 'answer' is K . If it is not expressly communicated, the probative value of the evidence is open to the risk of fallacious reasoning: '*Even if the expert's evidence is accurate and clear, there remains the challenge of successfully communicating the true probative potential of [DNA] evidence to the trial judge and to jurors. Lay jurors are likely to need some guidance in making sense of evidence expressed in terms of probabilities.*' [19] The Nuffield Foundation [35] make a direct reference to the importance of K : '*If the meaning and probative value of evidence depends on a background of contextualising information and the analytical framework used to interpret it – and it does – then the only way in which a lawyer or court can hope to grasp the meaning and probative value of scientific evidence is to acquire at least a rudimentary knowledge of the analytical framework employed by the scientist in arriving at her conclusions.*' This 'background of contextualising information' is the very foundation of K .

6.5 How is a prior formulated?

Since a prior is simply a probability of two competing hypotheses compared with each other, it is obvious that the hypotheses themselves must be considered in probabilistic terms before a prior is formulated. Of course there is a difference between finding a prior for $P(H)$ and finding a prior for $P(E|H)$, but value K is conditional on BOTH $P(H)$ and $P(E|H)$. While this sounds trite, it is often forgotten that all evidence begins with a prior, and the approach is the same whether the prior is being considered for $P(H)$ or $P(E|H)$.

The Nuffield Foundation, in its third practitioner guide [35] for the application of probability to criminal trials explains how the prior for a piece of crime scene ‘trace evidence’ is formed: *‘The probabilities for [the node] depend on the extent to which the characteristics, positioning, etc of the crime-scene mark are indicative of its association with the offence. For example, a footprint on the external sill of a second floor window is, all else equal, more likely to have been made by a burglar than a footprint left on the external front doormat (which could have been made by the postman, a door-to-door salesman, a house guest, etc). Assigning such subjective probabilities is necessarily a function of the forensic scientist’s knowledge, expertise and experience...’* In this example, $P(E|H_p)$ = the ‘probability of seeing the footprint if the suspect is a burglar’ may be, say, $\frac{1}{2}$; while the ‘probability of seeing the footprint if the suspect is a postman’ may be, say, $\frac{1}{1000}$.

What is important to note is that the probabilities are based on the expert witness’s knowledge of the world, as the Foundation explains: *‘Expert opinion testimony routinely rests on such subjective impressions, as is perfectly evident whenever an expert is asked to speculate how common some particular feature or characteristic of interest is (e.g. how many people have curved spines or walk with a limp), or how often he encounters that feature, say brittle bones or retinal damage, in his clinical practice, etc.* Once the evidence has been observed, the competing hypotheses are compared with each other in a ‘likelihood ratio’ (LR), $P(E|H_p)/P(E|H_d)$, which can be used to show the ‘probative value’ (ie the weight of the evidence towards one of the competing hypotheses) of the evidence.

The process of converting knowledge of the world into probabilities is not an easy one. The Foundation [30] explains the complexity of the task: *‘For each technique, the scientist is*

required, as a first step, to think through all the potential observations and analytical results that may be obtained were that technique to be employed. Then, probabilities for obtaining these outcomes are assigned given that the prosecution proposition ' H_P were true' and given relevant aspect of the case circumstances, i.e. the conditioning information. Next, the scientist has to adopt a completely different mind-set and assign probabilities for the outcomes if the alternative proposition ' H_D were true'... The process of assigning probabilities for the outcomes can be a complex one. The scientist will be combining knowledge and understanding from her own personal experience and from data and results contained in published experiments and surveys. There may well be numerous factors for the scientist to take into consideration, including the probabilities of transfer, persistence and detection, probabilities of obtaining 'matching' results, and probabilities of background occurrence and variability of materials. Different experts may arrive at different values of probabilities, reflecting their own experiences and knowledge of the evidence type as well reflecting different ways in which they have adjusted and used data from published studies. However, if the experts' probabilities were based solely on the same published data, then their probabilities should be very similar. The source and reliability of an expert's probability can, rightfully, be explored and challenged in court.' Therefore, while the probabilities are subjective and personal to the expert, and open to scrutiny in examination and cross-examination in court, it cannot be argued that the prior has not already been formulated by the expert before arriving at a probative value for the evidence. The probative value of the evidence posited by the expert is therefore directly related to the expert's own personal prior.

Let us take DNA evidence as an example. As the Nuffield Foundation [30] state: '*[In a case of DNA analysis], the proposition pair then being considered would be of the form:*

H_p : The DNA came from the suspect

H_d : DNA came from an unknown person, unrelated to the suspect

Assume that, on analysis, a full profile matching the suspect was obtained. The probability of obtaining this match, conditioned on the prosecution proposition, would be approaching a value of 1, i.e. it is practically certain that a match would be obtained. On the other hand, the probability of obtaining this match, assuming the alternative proposition were true, would be of the order [of say] 1 in a billion. This gives a likelihood ratio of the order 1 billion and the

scientist could conclude that the matching profiles provided extremely strong support for the view that the DNA came from the suspect rather than from some other, unknown, unrelated person.' The interesting thing to note here is that the two competing hypotheses have already been formulated from the expert's knowledge of the world and the science of DNA testing, as the Nuffield Foundation [19] recognises: *'DNA profiles are assessed by reference to a pair of competing propositions formulated by the forensic scientist (or anybody else undertaking a similar evidentiary assessment), following established protocols and utilising her case-work experience and knowledge of the instant case.'* By comparing H_P with H_D , we know that this particular DNA sample is going to be presented to the jury as a RMP.

By comparing one person with other people in the world population, we have an implied assertion from the expert that the probability presented will take the form of a 'coincidental match' probability, and from this the jury can decide whether the probability is so low as to rationally rule out the probability that the defendant's DNA profile match was merely a freak coincidence. Redmayne [18] explains the theory: *'Once a match has been declared the expert must make a decision as to the significance of the match by calculating the match probability. Essentially, this involves assessing the probability of the match having occurred by chance. For this to be calculated, the scientist will need some knowledge of the frequency with which the alleles represented on the autoradiograph occur within a population'*. This knowledge of 'frequencies occurring in a population' forms the prior probability for a RMP.

The Nuffield Foundation [30] explains that DNA is not the only type of evidence which may be converted into probabilities: *'Pairs of propositions can be generated for each issue, taking into account the prosecution allegation and the assumed defence position. Some examples are given here:*

1. H_P – *This knife is the implement that was used to cut the washing line*
 H_D – *Some other implement was used*

2. H_P – *This is the shoe that made the mark in the garden at the time of the incident*
 H_D – *Some other shoe made the mark*

3. H_P – *Mr. S is the person who broke the dining room window at the time of the incident*
 H_D – *Mr. S did not break the window*

4. H_P – *Mr. S had sexual intercourse with the woman during the incident*

H_D – *Some other man had sexual intercourse with her*

In each of these proposition pairings, including the third if it is assumed that someone broke the window, the evidence is presented as a RMP – which is to say that the jury must decide whether the defendant is the suspect or whether his matching profile is merely a coincidence. In effect, the DNA evidence RMPs, and RMPs of other types of evidence, are materially the same - which is to say that when dealing with trait evidence and traces of those traits, all of these types of evidence contain the same basic priors with the same value K throughout.

6.6 Can the Likelihood Ratio (LR) be in any way ‘divorced’ from the prior?

K can only be preserved by ensuring that the LR is closely allied to the prior. It is important to show that the LR cannot be divorced from the prior, to show that different types of evidence cannot easily be combined without compromising K .

Traditionally, it has been thought that the probative value of a piece of evidence can be measured by how much the LR of that evidence changes the prior to a posterior probability.

In criminal trials, the closer to $P(E|H_P)$ the evidence is valued, the more probative effect towards the prosecution case it has. If the prior for a piece of DNA evidence, presented as a RMP, is $1/7P$ (which is to say that everyone in the ‘world population’ of around 7 billion people has an equal chance of being the suspect, before the evidence is presented) and the RMP is $1/1\text{billion}$, the posterior is $1/7$ – which is to say that the LR of the evidence is $[(P|H_P) = 1 / P(E|H_D) = 1/1\text{billion}] 1\text{billion}$:

$$[\text{Prior}] 1/7 \text{ billion} \times [\text{LR}] 1\text{billion} = [\text{Posterior}] 1/7$$

This makes intuitive sense, because after the evidence is presented, the probability of the defendant being the suspect is one in seven – the likely number of matches within the reference class. However, there is so much information involved in this fairly simple formula, that can it be said that the LR is not entirely based upon the prior probability? In other words,

can we formulate the LR without a prior probability, which, if not, means that it is not actually ‘independent’ of the prior?

Some key members of the mathematical community, including the Nuffield Foundation [30]: believe that priors are to be formulated without reference to the evidence in the case, which leads to a divorcing of the LR from the prior: *‘We stress that the function of assigning prior probabilities is subsumed within the jury’s fact-finding responsibilities and naturalistic reasoning; assigning prior probabilities should not be a matter for the expert, although they may be able to provide expert information that assists the triers of fact.’* How much ‘expert information’ is provided, will lead to a successful communication of K as long as the jury working assigning the probative value of the evidence has subsumed the expert’s prior entirely. If this has not taken place, the value K will be compromised and the probative value of the evidence will not be successfully evaluated. A simple example of this would be where during a brawl involving three people, one man is killed by a knife cut. One of the two surviving men is found to have blood on his clothes whose DNA matches the victim. If K is ignored in this case and instead the RMP is relied upon for $P(E|H)$ (where H is ‘defendant innocent of the knife attack’) then the answer will be completely wrong, because $P(E | H)$ is actually close to 1, since grappling with the victim after the knife cut would produce the transference of blood, regardless of which attacker had cut him.

Aitken and Taroni [38], supported by the same Foundation, confirm the position: *‘In criminal adjudication, the values of the prior odds and the posterior odds are matters for the judge and jury, in accordance with the normal division of labour in forensic fact-finding. The value of the likelihood ratio, however, is a matter for the forensic scientist or other expert witness, as it is an assessment of the objective probative value of their evidence.’* If we take this argument to its natural conclusion, it is plain that Aitken and Taroni favour a model which allows a LR to be formulated by an expert pre-trial, and then this LR is applied by a juror to his own prior. By combining a juror’s own prior – which is derived from evidence, not of the case, but of the world-at-large – this means that two different types of evidence are at risk of being combined, without sufficient heed to ensuring that K is preserved throughout the model as a whole. What the authors may be advocating here is that K is preserved for each individual piece of evidence. However, this will not allow evidence of different types to be combined in any meaningful way.

One of the main problems with an approach which might allow only discrete types of evidence to be combined and presented is that the jury may misinterpret the probative value of the evidence in relation to the case as a whole. This is understandable, since the jury's primary task is not to evaluate single pieces of evidence for its probativity towards a given hypothesis, but to return a verdict on the case based on the overall comparison of the ultimate H_p and H_d .

The Nuffield Foundation [16] acknowledge this primary task, but still do not explain how posterior probabilities towards the ultimate competing hypotheses of 'guilty' and 'not guilty' are formulated: *'Expert witnesses must not trespass on the province of the jury by commenting directly on the accused's guilt or innocence, and should generally confine their testimony to presenting the likelihood of their evidence under competing propositions. However, experts are not absolutely precluded from stating posterior probabilities relating to intermediate facts proving or constituting the offence, if invited to do so by the court and providing that such statements are appropriately qualified and contextualised.'* Even if the expert offers a suggestion of a posterior probability for the ultimate hypotheses, it is important to understand that this is not a 'trespass on the province of the jury' so much as a vital communication of K . Of course, the jury have the autonomous decision of accepting or rejecting the expert's weight of evidence, but this must be done through examination and cross-examination of the witness to allow the jury to scrutinise the values inherent in the model.

Confusingly, while the Nuffield Foundation [19] recognises the importance of the communication of the experts' views on the competing hypotheses which gave rise to the LR for a particular piece of evidence: *'Presentation is pivotal. Common sense tells us that the way in which evidence is presented to the fact-finder might be more or less conducive to its appropriate evaluation. Some forms of presentation may be relatively clear and informative, whilst others might be especially prone to misinterpretation or to confusing or misleading the fact-finder'*... the contradiction arises where the Foundation refuse to acknowledge the fact that the experts' views are effectively giving rise to a formulation of K . This is highly problematic for the unwary, because on the one hand it seems as if the mathematical community are adhering to key probabilistic principles by acknowledging and discussing their importance (albeit impliedly), but at the same time omitting putting these principles into practice.

There can be no clearer admission of this mistake than in the following passage in the Nuffield Foundation [35] guidance: *‘Bayes Theorem does not supply prior probabilities (except insofar as these are posterior probabilities which, in their turn, rest on prior probabilities which Bayes Theorem did not supply). The entire edifice, in other words, rests on subjective human judgements of “prior” probability – and the final calculations of probative value produced by Bayes nets can only be as good, or bad, as the initial human inputs.’* The Foundation’s assertion that BT does not supply prior probabilities is at the very heart of the problem. If the expert is not supplying prior probabilities to the jury when presenting evidence, the jury is left floundering, without *K* to protect it. Further, the whole concept of ‘probative value’ is founded on ‘proving a case in its entirety’ – something which cannot be done without a full Bayesian approach to the case.

This is a very strange state-of-affairs. On the one hand, the mathematical community seem very astute at designing probabilistic models (albeit without the use of a full BN) which evaluate complex cases with many competing hypotheses and many types of evidence, but on the other hand, there seems to be a grey area where the community have overlooked a basic tenet of probabilistic reasoning: ‘context’.

What makes the situation even stranger is that not only is the importance of context routinely mentioned, but it is seemingly rarely, if ever, applied. That such a fundamental mistake as this has been made by groups as experienced in probabilistic reasoning as the Nuffield Foundation [35], is almost beyond comprehension: *‘The probabilities for [the node] are the prior probabilities of guilt or innocence before taking account of [the evidence]. Of course, the forensic scientist is not entitled to make such determinations – these are questions for the jury – but this presents no difficulty, because what we are interested in, when modelling the probative value of scientific observations, is the likelihood ratio not the prior or posterior probabilities. We could stipulate any value for the prior probability of guilt without affecting the ratio of the likelihoods for the evidence. One value is as good as any other for these purposes.’* At one stage, the Foundation [30]: suggests that the expert’s prior might be ‘wrong’.

This is clearly very confusing for the unwary, who might assume that a prior could never be wrong (although it may come under scrutiny during the trial) if it derives from the expert’s

subjective beliefs: *‘In order to offer a posterior probability, the expert must have considered, consciously or otherwise, quantitatively or qualitatively, a value for the prior probability of the fact in issue. The risk here is that the expert’s prior probability may be inappropriate in the circumstances of the case: it may be purely speculative, it may not be informed by the case circumstances; it may be biased to one or other side of the argument’*. To understand precisely the problems which occur when different types of evidence are combined a practical study must be made. For this, it is probably easiest to take evidence such as DNA, which is generally presented in expressly probabilistic terms, and attempt to combine it with ‘other’ types of evidence to see the outcome.

6.7 DNA testing:

In DNA testing, where the identity of the suspect is a principal issue, the RMP is generally used to establish the probative value of the evidence. The RMP compares the following two competing hypotheses, which must be a true negation of each other, and therefore mutually exclusive and exhaustive [10]:

Hp: ‘That the defendant is the source of the evidence’

Hd: ‘That someone else in the world population, alive at the time of the crime, is the source of the evidence’

We know that Hp and Hd are mutually exclusive and exhaustive, because *someone* in the world population must have left the crime scene trace, and this hypothesis pairing covers all possible eventualities. By comparing $P(Hp)$ with $P(Hd)$ we have our prior. Since $P(Hp) = 1$ and $P(Hd) =$ the world population (‘WP’) - taken to currently be approximately 7 billion people - we can say that the prior $P(Hp)/P(Hd)$ for all DNA evidence presented as a RMP is $1/WP/1-[1/WP]$.

This is not controversial, and has been endorsed by Buckleton et al [6], who suggest that *‘true profile frequencies will take the values $n/\text{world population}$ ’*. Aitken and Taroni also endorse this approach: *‘As a starting point we suppose, simplistically, that before any evidence is heard, ‘innocent until proven guilty’ means that every person in the relevant population is equally likely to be guilty. If the relevant population were taken to be the population of the*

whole world it is fairly straightforward to think of evidence that will eliminate most of the people in the world from serious consideration as potential suspects.' The legal community should also be satisfied that the $1/WP$ prior satisfies the 'innocent until proven guilty' principle, since before evidence is heard, the defendant has a uniform probability of equal guilt with that of the rest of the world population. It is simple to check the accuracy of this prior by applying BT:

Hp: 'That the defendant is the source of E'

Hd: 'That someone else in the world population, alive at the time of the crime, is the source of E'

E: DNA with a RMP of 1/1million

$1/WP$ (prior) \times 1 million (where $P(E|Hp)$ is taken as 1, if we assume a 0 error probability rate) = 1/7,000 (posterior).

The posterior probability reflects our belief in $P(Hp)$ as 1/7,000, which is consistent with there being 7,000 people in the world population who share identical DNA evidential profile traits. Therefore Tribe's argument that it would be difficult for a juror to formulate an accurate prior is without foundation.

Finkelstein and Fairley [23] argue that the prior should not be too large for fear that inferences drawn from it would be inaccurate. This argument also has little foundation, as it can be seen here that the ONLY rational prior for DNA presented as a RMP, is $1/WP$. Of course, this value is only accurate if value K , which is at the foundation of this reasoning, is communicated to, and accepted, by the juror. Aitken and Taroni [38] have confirmed the approach as follows: *'[W]e only require prior assessment of relative measures of the different propositions of the defence. These [a]re called "prior ratios", and in the context of forensic DNA investigation one of the prior ratios was interpreted as the number of times more likely the "average" full sibling (of the siblings considered) is of being the donor of the DNA, than the "average" unrelated individual (of the unrelated considered), before consideration of the forensic DNA evidence. The assessment of these prior ratios, along with the number of objects or individuals pointed out by each of the defence propositions, enables the value of evidence to be expressed as a single formula. Thus, we try to consider the complete set of alternative donors of the DNA (both close relatives and unrelated individuals to the suspect)*

and how the differing match probabilities between these may affect the case against the suspect.’ In fact, from this assessment of the prior, leading to a RMP (and then an LR, of course) of a piece of DNA match evidence, we can even go so far as to make the following key RMP evidence proposition:

‘A RMP treated by the probabilistic model reflects the overall number of suspects who share an equal probability of guilt’.

One further issue which must be discussed before moving forward, is the assumption that $P(E|H_p) = 1$. In other words that where the DNA trace has been found at the crime scene, it is assumed that the probability of seeing the trace if the prosecution hypothesis (‘that the defendant is the source of the DNA trace’) is true, = 1, or 100% certainty. This is a huge assumption to make and is not realistic. In their article, ‘*Subjectivity and bias in forensic DNA mixture interpretation*’, Itiel E. Dror and Greg Hampikian [55] discuss the results of empirical research which suggests that DNA mixture interpretation is subjective and can produce opposite conclusions amongst qualified analysts on the basis of extraneous context.

Dror and Hampikian recognise a long line of research which has established that seeking and interpreting information based on a set of existing beliefs can produce biased results. In the field of fingerprinting, the same forensic experts may arrive at different conclusions when identical evidence is presented within different extraneous contexts. DNA, in contrast to other forensic disciplines, has come to be regarded as the gold standard of forensic science and held to be objective and immune to subjectivity bias. In fact DNA has been elevated to the status of a ‘laboratory-based science’ along with toxicology and drug analysis, as apart from other disciplines which are ‘based on expert interpretations of observed patterns – fingerprints, writing samples, bite marks etc’.

If correct, then DNA is not affected by contextual circumstances, but in complex situation – such as DNA mixtures, there is the argument that DNA analysis may be subjective and relies on a number of factors. This objective/subjective debate had not been properly researched until Dror and Hampikian’s test:

6.7.1 Dror and Hampikian's DNA analysis test:

1. The study was conducted with qualified DNA experts who conduct real casework in accredited laboratories.
2. That the examiners genuinely believed the contextual framework, since contrived contexts would not give the required effect or impact.
3. Mixture DNA used from a real adjudicated criminal case, with DNA profiles of the victim and three suspects, chosen to provide analysis within an extraneous context.

(The DNA evidence related to a gang rape case in which one of the assailants testified against other suspects in return for a lesser sentence as part of his cooperation in a plea bargain deal. However, those identified through the plea bargain denied any involvement in the rape. The DNA conclusions were critical to prosecution. If the suspects were excluded by DNA, or even if the DNA was 'inconclusive', the incriminating testimony of the admitted rapist would most likely not be allowed. As potentially biasing as this domain irrelevant context was, it should not have affected their analysis. The mixture DNA from the sexual assault was examined by experts in the initial criminal case, and their analysis and conclusions were that the suspects that were identified by the cooperative assailant could not be excluded from being contributors to the mixture.)

4. The mixture was presented to 17 examiners, fourteen female/three male with a mean age of 41 and a mean experience conducting DNA analysis of 9 years, without the contextual case information.
5. The results within those 17 examiners were compared to test the subjectivity in DNA analysis to assess subjectivity in DNA analysis and then compared with results from those who examined the mixture, within the contextual framework, to assess bias in DNA analysis.
6. Each of the 17 DNA examiners independently examined the evidence, and gave one of three conclusions for each of the suspects: 'cannot be excluded', 'excluded', or 'inconclusive.' One of the suspects, suspect 3, was the point of interest, as he was determined

as ‘cannot be excluded’ by the DNA examiners who examined his DNA within the potentially biasing context.

In regard to suspect 3, the results from the 17 examiners varied. One examiner concluded that the suspect ‘cannot be excluded’, 4 examiners concluded ‘inconclusive’, and 12 examiners concluded ‘exclude’; which suggests subjectivity in DNA interpretation. If it was objective, then since the examiners all work in the same laboratory and work under the same interpretation guidelines, the answers would all have been the same. The subjectivity may reflect differences in training, experience, personality and motivation.

Comparing the data between examiners from the context free environment with those who were exposed to the extraneous content framework: The DNA analysts who concluded that the suspect cannot be excluded within the biasing context of the criminal case, are in sharp contrast to the vast majority of examiners who examined the same evidence without this biasing context. Only 1 (out of 17) gave the same conclusion as the original analysts, 16 other examiners reached a different and conflicting conclusion (either ‘exclude’, 14 examiners, or ‘inconclusive’, 4 examiners). Thus, the extraneous context appears to have influenced the interpretation of the DNA mixture. It must be emphasized, however, that these effects were observed for a DNA mixture analysis. Previous research in forensic identification suggests that contextual influences are most powerful when the evidence is ambiguous, complex, and a ‘hard call’.

6.7.2 Low template DNA (LTDNA)

David Balding [56], an expert in statistical analysis of DNA profiles, has stated that where there are very few (typically three, approximately) sample cells to examine (known as ‘low template DNA’ or ‘LTDNA’), that there is a general danger of material ‘dropping-out,’ whereby alleles simply disappear from the sampling data, or potentially more damagingly, ‘dropping-in’, whereby a molecule of someone else’s DNA contaminates the sample, even in what has been considered a ‘clean’ environment. Considering the probability of whether the same could be said of higher copy DNA samples, is really a matter of common sense. If vital cells can simply disappear or appear, despite the best efforts of the expert, there must be an even higher probability of contamination occurring anywhere else along the chain, where the

persons handling the sample may not be quite so careful. Even if one accepts that the examination process is a perfect science, the variables in this process which contain elements of uncertainty may include a misreading or misinterpretation of the data by the expert (as discussed by Dror and Hampikian above), contamination of the sample, accidental swapping of the data with a different sample in the laboratory, and countless other opportunities for error. None of these variables can be said to be extrapolated from a firm statistical base, which is what makes DNA sampling as uncertain as any other form of evidence.

Therefore the assumption that $P(E|H_p) = 1$, is based on a huge and irrational leap of faith that we can be 100% certain of anything. For the sake of this discussion, we will allow '1' as the basis for $P(E|H_p)$, but in reality a BN would be necessary to model the inevitable uncertainties which would arise and would certainly bring the probability value to a figure less than 1.

DNA is not the only type of evidence which is capable of being presented as a RMP. Blood type, hair colour, or fingerprints which proliferate randomly in the world population are all capable of being presented as a RMP. We call this type of evidence 'trace' evidence, because it pertains to an unknown suspect who is said to be the source of the evidential trace. Even an eyewitness giving evidence that the suspect had a particular physical trait – such as skin colour, height, distinguishing facial features etc - can be considered 'trace' evidence, and that trait should be presented as a RMP.

In the case of $R \vee T$ [2], an expert witness used BT to calculate the probative value of rare footwear marks, but this was rejected by the Court of Appeal as footwear marks were said not to derive from a '*sufficiently certain database*'. However, this ruling has been criticised for its irrational reasoning [47]; which means that theoretically any evidence which has its probative value calculated by dividing its likelihood of a match with the defendant's profile by a random or 'coincidental' match in the world population, is capable of being presented as a RMP. What is interesting about this is that it means that different pieces of evidence can successfully be combined to create a single posterior probability. What is necessary to achieve this is a consistent value K throughout the model.

6.8 Combining evidence with different RMPs:

Let us say that there are two completely independent pieces of evidence in a case: A sample of DNA left at a crime scene with a RMP of 1/ 1million, and a rare footwear mark with a RMP of 1/100. The case is as follows:

Hp: 'That the defendant is the source of E^1 , E^2

Hd: 'That someone else in the world population, alive at the time of the crime, is the source of E^1 , E^2

E^1 : DNA with a RMP of 1/1 million

E^2 : Footwear mark with a RMP of 1/100

By combining our 1/WP prior with E^1 , we calculate our posterior as 1/7,000 (of course, assuming E^1 and E^2 are independent). This posterior can now be used as the prior for our new piece of evidence:

$1/7,000$ (prior) $\times 100$ (where $P(E|Hp)$ is taken as 1 if we assume a 0 error probability rate) = $1/70$ (posterior).

What this means is that in our pool of 7,000 random suspects, we would expect to find 70 suspects who carry both evidential profiles. In effect, the 70 suspects have been filtered twice by E^1 and E^2 from the same pool of suspects WP, making E^2 a 'sub-class' of E^1 . This makes intuitive sense. There would be 70 'coincidental' matches from the original 1/WP prior.

We could equally say that E^1 and E^2 has a combined RMP of 1/100 million, using the product rule - much in the same way that DNA loci are multiplied together to give a master RMP for the entire sample. The order in which these two pieces of evidence is presented is immaterial, which means that Tribe's [4] second objection, that a large prior would overshadow other types of evidence, is also without foundation. This 'anchoring bias' [57] a natural human trait which is argued to be prevalent in society is actually *eradicated* by a transparent and properly communicated application of BT.

6.9 Combining RMP evidence with ‘propensity’ evidence:

If we accept that we can combine evidence RMPs, can we combine an RMP with other types of evidence *not* presented as a RMP?

This following passage shows that the Nuffield Foundation believe RMPs can ‘relate’ to other evidence in the case: *‘Under the Doheny and Adams approach, random match probabilities displace likelihood ratios in courtroom testimony... If members of the public are, generally speaking, more familiar and comfortable with probabilities than with numerical likelihood ratios, it is reasonable to suppose that jurors will be better equipped to assess the probative value of DNA profiling evidence expressed as an RMP; especially if trial judges further spell out the logical implications of the RMP, as the Court of Appeal encouraged them to do in Doheny and Adams... However, this still assumes that jurors can make sense of the RMP, as a quantified measure of probative value, and relate it to the other evidence in the case.’* For this reason, it makes no sense to give LR_s for individual pieces of evidence. An LR might be acceptable only for the ultimate ‘guilt’ v ‘innocence’ hypotheses, but for all evidence in a case a BN is the only means of fully explaining the interdependencies.

Since the case of Roy Whiting [58] and the coming into force of the Criminal Justice Act 2003, which widened the gateways to allowing defendants’ bad character evidence into his current trial, there has been a growing interest in combining propensity evidence with other types of evidence. The question is, is it possible to combine DNA with propensity evidence? Let us take the following case example:

Hp¹: ‘That the defendant is the source of E¹’

Hd¹: ‘That someone else in the world population, alive at the time of the crime, is the source of E¹’

E¹: DNA with a RMP of 1/1 million

Hp²: ‘That the defendant is guilty’

Hd²: ‘That the defendant is not guilty’

E²: The defendant’s criminal record showing a strong propensity to commit similar crimes

After observing E^1 : $P(Hp|E^1)/P(Hd|E^1) = 1/7,000$

Let us say that the juror ascribes a LR of 2 for E^2 , since it is his belief that the probability of seeing the propensity evidence is twice as likely if the defendant is guilty than if he is innocent. Therefore, $1/7,000$ (prior) $\times 2$ (LR) = $1/3,500$ (posterior).

At first glance this makes intuitive sense. The juror's posterior reflects an increase in the belief of the defendant's guilt after observing E^2 . However, this increase is without foundation. Taking our aforementioned key proposition; if we accept that our prior of $1/7,000$ reflects the number of equally guilty suspects then our posterior should do the same. E^2 has effectively halved the number in our pool of suspects, but this cannot be accurate.

Just because the defendant has propensity evidence, which has prosecution probative effect, does not mean that the number of suspects in our pool should reduce. We have not tested the other 6,999 suspects for propensity evidence of their own. Theoretically all 6,999 suspects, if they could be found and investigated, could *all* have propensity evidence against them - which means that the number of suspects cannot be reduced on the strength of E^2 .

To clarify: Since we have not been able to establish that E^2 is a sub-class of E^1 , we **MUST** assume that they are independent of each other, and we cannot draw any inferences about connections between the two. E^1 and E^2 are both important pieces of evidence in the case, but they have not been properly combined in a single model. Mike Redmayne [59] identifies the inherently different types of evidence, and potential difficulties in combining them, in the case of *R v Adams*, when he compares a '*type of alibi*' (implying that the defendant was not at the scene of the crime and could not have left the trace) with the '*sort of person who would have an alibi*' (arguing that the defendant's character is such that his alibi is credible).

Unfortunately, it has not yet been established that these two types of evidence, while maybe important as pieces of evidence in their own right, should not be used in combination to provide an overall probability of guilt. Aitken and Taroni [38] explained how in *R v Adams* the evidence was combined: '*In Adams it was necessary to assign numerical values, not only to the explicitly probabilistic DNA evidence, but to all pertinent information before the jury. For example, the victim gave a description of her attacker which was hard to reconcile with*

the defendant. She also failed to pick out Adams at an identity parade. For the purposes of illustration, a probability of 0.1 was assigned to this (palpably weak) identification evidence on the assumption that Adams was guilty, as against a probability of 0.9 assuming his innocence. This gives a 'likelihood ratio' of 9 in support of innocence. In addition, a former girlfriend of Adams gave an alibi which was not challenged at trial. This evidence was assigned a probability of 0.25 if the defendant were guilty and a probability of 0.50 if he were innocent. The jury were then instructed in how to combine these probabilities with 'prior odds' of guilt of 200,000 to 1 against and DNA evidence of 1 in 20 million (as calculated by the defence).' As can be seen here, the figures a must be compelling to an unwary juror - small probabilities being combined with tiny ones to provide a minuscule probability of guilt – but as has been seen here, these combinations are not only illegitimate, they mean absolutely nothing. Of course, in Adams, the key problem was that the hypothesis of the defendant being 'guilty' was treated as being synonymous with him being the 'source of the DNA evidence' – which are two very different things.

To illustrate this point, Professor Norman Fenton [60], an expert in creating BNs to show causal pathways in complex criminal cases, proposed the following model - before problems with the model were highlighted in this thesis - in cases with similar evidence as Adams:

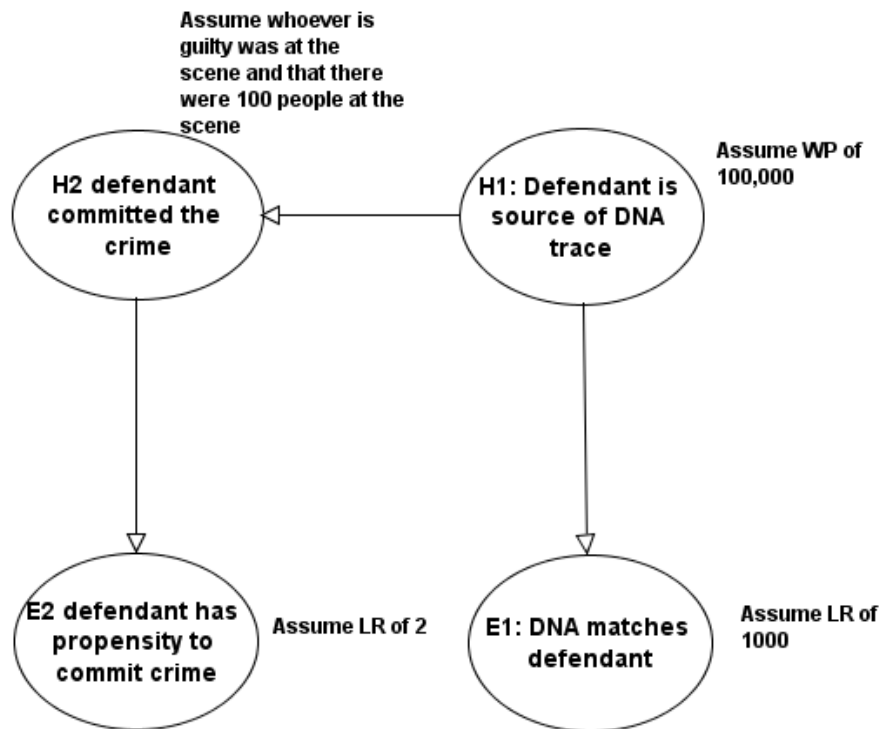
H1 = The defendant is one of the world population who, before evidence is observed, is in the pool of suspects in the world population, with equal probability of being the source of a DNA trace.

H2 = The defendant is one of 100 people seen at the scene of the alleged crime, with equal probability of having been the attacker.

E1 = A trace of DNA found at the scene of the alleged crime, which an expert states has a random match probability (RMP) of matching 1 in every people in the world population.

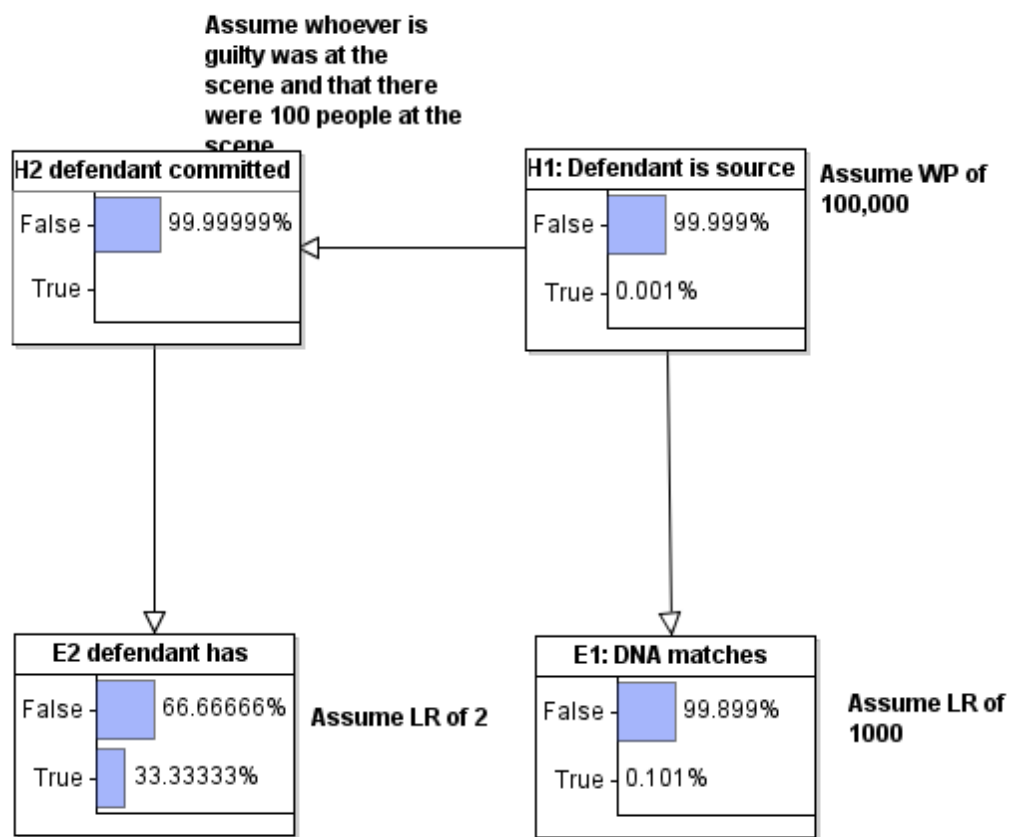
E2 = Evidence of the defendant's past criminal action, which an expert states is twice as likely to be observed if the defendant is guilty of a similar crime (such as the one that he is currently charged with) than not.

Fig 6.9(a)



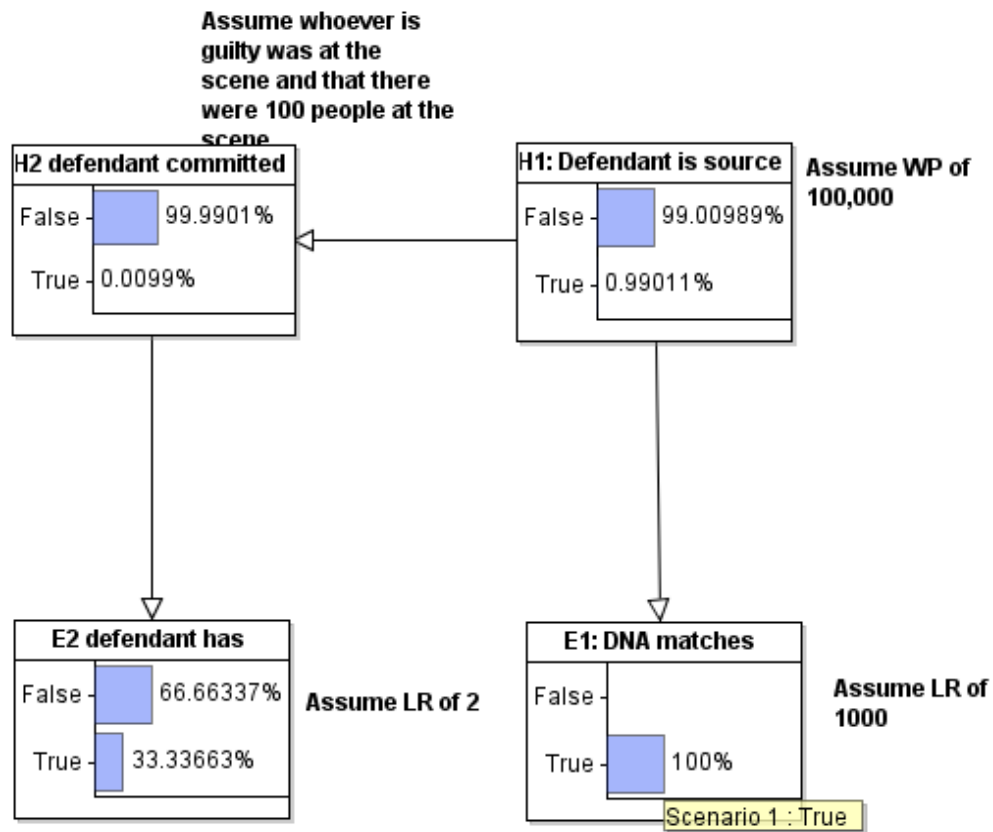
.and below is this causal model with the priors added:

Fig 6.9(b)



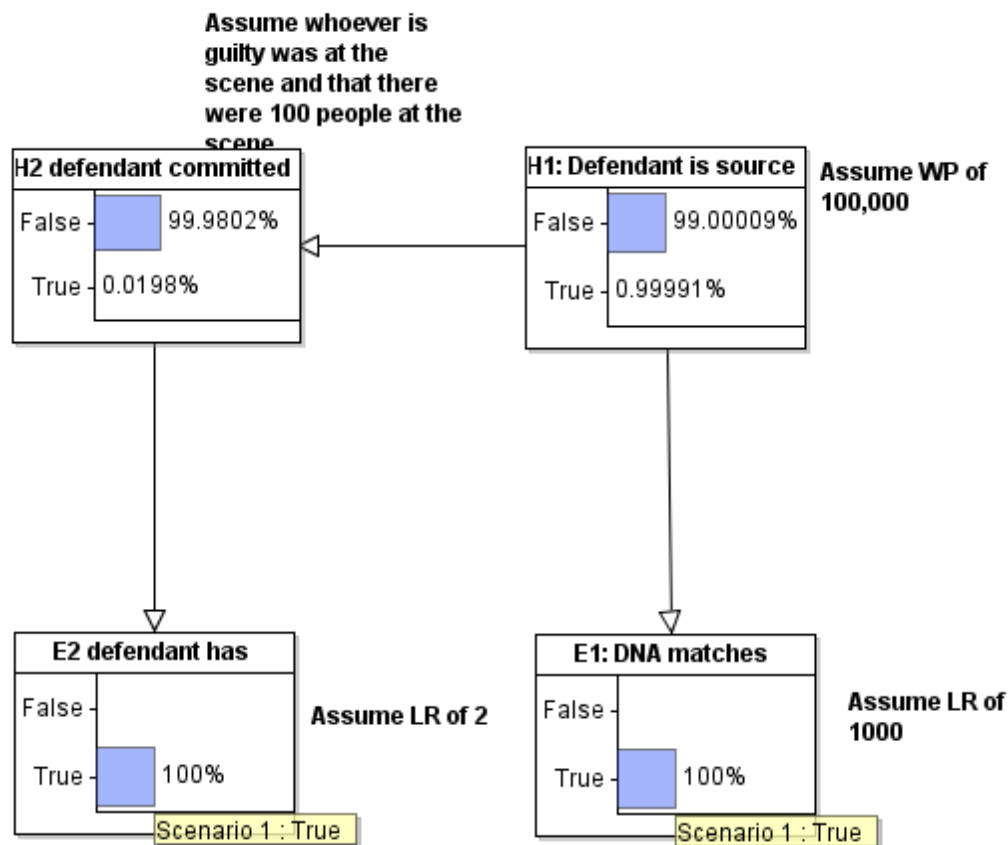
...and with the DNA evidence E1 added:

Fig 6.9(c)



...and with E2 added:

Fig 6.9(d)



As can be seen with this model, Professor Fenton, who, under standard practice, has attributed H1 with a prior of 1/100,000 to reflect a hypothetical world population, together with a H2 prior of 1/100 to reflect ‘location evidence’ of 100 people observed at the scene of the alleged crime, has then combined with E1 - a DNA RMP of 1/1000 (giving a LR of 1000, ie $P(E|H) = 1$ divided by $P(E|\neg H) = 1/1000$) – and E2 – propensity evidence with a LR of 2 (ie $P(E|H) = 1$ divided by $P(E|\neg H) = 1/2$), to provide an overall posterior probability of guilt.

While this may seem intuitively correct, due to smaller and smaller probabilities of innocence being generated as more and more evidence and hypotheses are added to the model, these probabilities have no foundation in reality, nor do they respect the jury’s task of attributing weight to evidence, for the following reasons:

1. If we accept the H1 prior of 1/100,000 for the hypothetical entire world population, this prior merely asserts that the attacker must have come from the world population, and this in turn establishes that someone must have committed the crime and that before any other evidence is observed, that everyone in the population has an equal probability of being the attacker;
2. That H2 places the attacker into a smaller sub-class of the world population than 100,000. While it is tempting to assume that H1 and H2 can be combined, they actually conflict. There is no way that we can create a smaller probability by combining 1/100 and 1/100,000, because to do that we would have to accept that the defendant being at the scene is fact rather than evidence. In effect, where there are two completely different and conflicting reference classes, which of these two classes do we prefer?
3. That to bring E1 into the model, we would have to assume that DNA match evidence creates a 'force field' around each potential match, keeping them homogeneously spread within the world population (ie spread evenly through the sample space) and perfectly evenly spaced from each other, in order that we could capture a precise number from within the sub-class; and
4. That to bring E2 into the model, we would have to assume that the attacker is the only person in the sub-class who has the adverse propensity evidence weighing against them. In other words, the probability of innocence only becomes smaller here because we are assuming that everyone else in the sub-class is 100% certain not to have propensity evidence – not only absurd, but also asserted as fact rather than evidence.

For these reasons, the posterior probability of guilt of the defendant having committed the crime, of 99.802% (H2), is staggeringly high, and would almost certainly convince a jury to convict, while not having any foundation in reality. Of course, this high 'guilt' probability greatly increases the risk of a miscarriage of justice due to an overvaluation of the prosecution evidence in the case.

To expand on these issues, 'location evidence', 'homogeneity' and problems with assumptions of 'perfectly evenly spaced' DNA matches are investigated and discussed below:

6.10 Combining RMP evidence with location evidence:

‘Location evidence’ places the defendant at or near the scene of the crime. It differs from trace evidence, in that the defendant himself, rather than just some of his physical traits, is known. Redmayne [18] explains that it is common practice in criminal trials to combine trait evidence, such as DNA, with location evidence: *‘It has been shown that the jury plays a key role in combining DNA evidence with other evidence adduced at trial, in particular by making decisions about the suspect population.’* Location evidence can come from a database of people living in a particular area, of which the defendant is a member; CCTV evidence placing the defendant at the scene of the crime; a witness who knows the defendant and saw him at the scene; or even an admission from the defendant himself that he was there. By placing the defendant at the scene of the crime, we can start to calculate the probability of his guilt.

Let us say that the defendant was at the scene of the crime with the victim and one other unknown person and we know that there are no other suspects. We can therefore say that the probability of the defendant’s guilt is 50%. This must be so, because in the particular facts of the case it is assumed that the victim could not have killed himself, and the other accused person had an equal chance with the defendant of committing the crime. In a mob of 50 people surrounding a victim, of which we know that the defendant was a member, we can say that his probability of guilt is 2%. We also know that the individual probability of guilt for everyone else in the mob is also 2%.

The key to location evidence is that, due to the ‘mutual exclusivity/exhaustivity’ principle [10] the probability of guilt is equal with the size of the sample space, which means that the larger the pool of suspects, the smaller the probability of the defendant’s guilt. In this way it does not matter how close the defendant was to the scene of the crime, because location evidence designates a finite pool of suspects to be afforded an equal probability of guilt. It must also be noted that location evidence does not require the identities of the other suspects in the pool to be known in order that the probability of the defendant’s guilt is calculated.

In Adams [54], the location evidence was ‘153,000 men in the Hemel Hempstead area’ - which provides a probability of guilt of $1/153,000$ for each of this pool of suspects, including the defendant who was stated to live there - while in Collins [5], the location evidence was

‘couples in the Los Angeles area’ – which provides a probability of guilt of $1/n$ for each individual couple, including the Collins’ who were said to reside there, where ‘n’ is the sample space derived from a census database.

In the U.S. rape case of *People v Puckett* [22], the defendant’s DNA profile was checked after discovering he had a criminal record in the state of California for similar offences. Colmez and Schneps [20] argue that the fact that the defendant came from California, provides a prior 0.12% probability of guilt based on the size of the population of that particular region. Redmayne [18] also endorses this approach: *‘In cases where there is no evidence against a suspect apart from DNA evidence - and with the introduction of offender databases such cases may become more common - the prior odds, which cannot be zero, must be based on the “suspect population”, the number of people who could possibly have committed the offence. The prior odds may then reflect the number of people in a city or even in a country. The prior odds of 0.000001 used in the example above would correspond to the jury’s belief that the perpetrator of the crime could be anyone in a city with a population of one million.’* The suspect population evidence is combined with the DNA evidence to provide an overall posterior probability of guilt. In effect, the suspect population, or ‘location evidence’ or ‘reference class’, becomes the prior, which is then to be combined with the LR of the DNA evidence in a Bayesian model.

Fenton and Neil [3] suggest that location evidence can formulate a prior. They argue that evidence of a mob of 100 people surrounding the victim, of which the defendant was a member, can form a prior of the defendant’s guilt of 0.01 to be combined with other evidence.

In fact, Fenton and Neil also argue that the jury should be allowed to select a ‘range prior’ or a single point prior from that range, to suit its own belief of the accuracy of the evidence, thereby preserving its autonomous, fact-finding role. In Fenton and Neil’s example they combine this prior of 0.01 with a piece of DNA match evidence found at the scene of the crime, which has a RMP of 1/1 million, thereby calculating a posterior probability of guilt of 99.99%, or in other words that *‘the probability of innocence is now one in 10,000.’* This case example is therefore:

Hp¹: ‘That the defendant is the source of E¹’

Hd¹: ‘That someone else in the world population, alive at the time of the crime, is the source of E¹’

E¹: DNA with a RMP of 1/1 million

Hp²: ‘That the defendant is guilty’

Hd²: ‘That the defendant is not guilty’

E²: A mob of 100 people, of which the defendant is a member, surrounding the victim

Taking 0.01 (E²) as our prior, we combine it with the LR of the DNA evidence $P(E|Hd^1)/P(E|Hd^1) = 1 \text{ million } (E^1)$ to calculate a posterior of 10,000 in favour of the prosecution hypothesis, or as Fenton and Neil say: ‘*the probability of innocence is now one in 10,000.*’ The question is what does this posterior actually mean? If we take our RMP key proposition, the posterior should reflect the number of suspects still in the fray. In this example the number of suspects is 1, since the posterior has breached the WP threshold (to clarify, if the posterior had been 1/10,000, the number of suspects would have been 10,000, but in this case the posterior is 10,000/1, which means that there is only one suspect: The defendant – as long as you are able to confirm the DNA of the other people who were at the crime scene does not match). The problem, however, is that we have lost the value K throughout the model. The 0.01 prior is not a RMP, which means that E¹ and E² do not share common values. In other words these two types of evidence are incompatible. Colmez and Schneps [20] similarly argue that a 0.12% probability of guilt based on location evidence, in the case of Puckett [22], can generally be combined with RMPs. As has been shown here, this is simply not the case.

It might be argued that regardless of the disparity in K between E¹ and E², that the inferences drawn are still relevant. It might be said that a posterior of 10,000 accurately reflects the juror’s beliefs in the probability of the defendant’s guilt – which is precisely the reason the defendant is in court - regardless of whether the prosecution hypothesis should be Hp¹ or Hp². However, to adequately explain why this is not the case, an examination of each element of the case must be made:

Taking E¹ firstly: The hypothesis pairing of Hp¹ and Hd¹ leads us to a prior for this piece of evidence of 1/WP - which means that the posterior of this evidence alone is 1/7,000, meaning that there is a pool of 7000 suspects before E¹ is combined E².

Taking E^2 secondly: The hypothesis pairing of H_p^2 and H_d^2 leads us to a prior for this piece of evidence of 50/50 or 1 - which reflects the division of two equal and opposite mutually exclusive and exhaustive [10] states. The posterior for this evidence is 1/100, meaning that there are 100 suspects before E^2 is combined with E^1 .

The question is, how can E^2 reduce the number of suspects from 7,000 to 1? Well, much in the same way that propensity evidence cannot reduce the number of suspects in E^1 , nor can the location evidence, for much the same reason: *The rest of the sample space has not been tested*, which means that we have no way of knowing how many of the 7,000 suspects are also in the mob of 100 people. In essence, E^2 is not a sub-class of E^1 . Tribe [4] called this lack of knowledge of a particular probability a ‘soft variable’, in which if it cannot be counted it cannot be said to exist. However in this case, since the 100 suspects have not been counted, rather than not able to be counted, the same argument can be made.

Under circumstances where K is consistent for both E^1 and E^2 , the evidence can be combined. Taking E^2 in this example, where the posterior for this evidence is 1/100, this would mean that there are 100 suspects each with an equal probability of supporting the prosecution hypothesis of guilt. If the prior for this evidence was 1/WP as for E^1 , it would mean that there would be 100 suspects randomly placed within the world population.

However, to reach this posterior, the LR for this evidence would have to be WP/100 – a very large number, and almost certainly not one in the contemplation of the witness (or any other means of arriving at the posterior) when arriving at 1/100. Even if this were so, the difference between 1/100 as a sub-class of 1/WP and 1/100 as an arbitrary probability is crucial. Combining evidence with a LR of 1 million (E^1) and WP/100 (E^2) from a 1/WP prior would give a posterior of approximately 1/70.

This posterior reflects the filtration process of two pieces of evidence with the commensurate coincidences necessary to arrive at small probabilities. In essence, a rarefied 100 suspects from a world population of 7 billion people means that they all share a very special trait not seen outside of the group. But with an arbitrary 1/100 probability based on 100 suspects picked at random without the necessary filtration process? The coincidence factor is not there, which means that the small probabilities cannot be justified.

What makes this ‘lack of coincidence’ problem even more acute is that the 100 suspects were not individually plucked at random from around the world population and then placed in a group to be evaluated. They were picked in a single group located around the victim. This factor means that there is the added problem of secondary transfer of E^1 to E^2 by a number of different methods, including innocent transfer due to natural excretion, negligent transfer by emergency and investigating staff, or deliberate transfer by the guilty party seeking to contaminate evidence, simply because the suspects were in such close geographical proximity to the victim.

Paradoxically, in this case $P(E|Hd^2) \approx 1$, since the probability of seeing the evidence would be close to certain (assuming a 0 probability error rate, of course) if the defendant was innocent, which means that the LR - $P(E|Hp^2)/P(E|Hd^2) - \approx 1$ also. Strictly speaking this would mean that the evidence is ‘neutral’ [10], which causes further problems in ascertaining its probative value.

On the other hand, it might be argued that because the prior for any RMP of DNA is $1/WP$ - which means that the evidence proliferates randomly around the sample space, thereby evenly spacing out the suspects - that whoever is closest to the victim is most likely to have committed the crime. The proposition would be that since E^2 places the defendant very close to the victim, he has a higher probability of being the attacker than anyone else sharing the DNA traits.

If we take a very recent example of this way of thinking, we notice a flaw: *‘Prior probabilities are, as in any Bayesian approach, particularly difficult to elicit. In our network, the priors that need to be specified are the numbers in the probability tables for the scenario nodes. They represent the prior probability of a scenario, which could arguably be viewed as the plausibility of a scenario as it occurs in narrative research, expressing how likely a person would find a scenario beforehand. One can really only subjectively estimate these probabilities. Any attempt to objectively estimate this plausibility gives rise to a number of issues. Firstly, there is the principle that in court no prejudice should be held against any suspect, which is sometimes argued to mean that prior probabilities should be equal for everyone. However, this does not imply that two scenarios about different suspects should always have equal prior probabilities: that also depends on the coherence of the scenario.*

Secondly, there is the reference class problem concerning how much detail to include in the prior probability. Knowing that a suspect lived in the neighbourhood where the crime took place undoubtedly leads to a higher prior probability (because there was more of an opportunity) while knowing that a suspect lived on the other side of the world decreases the prior probability. But even when the prior probability does not take into account in what part of the city the suspect lived, there will always be some prior information included, such as that the suspect lived somewhere in the country, on the continent or somewhere in the world. A subjective estimate of the priors still runs into the principle of equality in court: no distinction should be made based on prejudice. Therefore, in our model of the Anjum case, both prior probabilities for the scenario nodes being true were set to 0.001. Another problem with a subjective estimate, however, is that once the model is done, one might be tempted to view these numbers as if they are objective. In other words: explicitly quantifying the priors might lead to the false illusion of objectivity. We strongly emphasize that our method is meant to formalize subjective accounts of scenarios. In this perspective, our method supplies techniques to support a subjective decision with more formal tools.’ [61] The crucial flaw is to be found in the following assertion: ‘Knowing that a suspect lived in the neighbourhood where the crime took place undoubtedly leads to a higher prior probability (because there was more of an opportunity) while knowing that a suspect lived on the other side of the world decreases the prior probability.’ While this may be true intuitively, the real question is how this may be presented probabilistically.

If we refer back to the ‘key RMP evidence proposition’ made earlier, the only question needing answering by the jury is ‘How many suspects are there?’ From the example above we start with a prior probability of 0.001, which means that there must be 1000 suspects before any other evidence is observed. The jury will know this from the following implied assertion: ‘that the suspect lived somewhere in the country, on the continent or somewhere in the world.’ After the other evidence in the case is combined, the jury will want to know exactly how many suspects there now are. Can this be done with any degree of accuracy? To answer this question, one must examine the assertions and their composite parts.

Firstly, there are assumptions as to the sample space. If we must assume that those living closer to the victim have a higher probability of committing the crime, we must assume homogeneity.

6.10.1 Homogeneity:

Homogeneity in the world population does not mean ‘perfectly evenly spaced’. In fact, Buckleton et al [6] state that it would be an error to assume homogeneity from one sample group to another. The danger of the CPS stating that ‘the old woman from Beijing’ would not be of interest to the investigation of a rape in Huddersfield, is that it implies that she would be too far away from the crime to be culpable. In any event, even if homogeneity was assumed to mean ‘perfectly evenly spaced’, this does not explain why there would be an element of uncertainty in the probability of a defendant sharing DNA traits with that of a crime scene stain presented as a RMP of, say, 1/1billion. As there would only be seven suspects in the world population and the victim and the defendant are found to be located together in a group of 100 people, why would there be any uncertainty at all that the defendant was the attacker?

If the ‘perfectly evenly spaced’ homogeneity principle held true, the probability of the attacker being any of the other 6 suspects would be nil, which of course, in reality, is absurd. In Fenton and Neil’s example, *‘the probability of innocence is now one in 10,000’* has no logical basis at all, while Colmez and Schneps [20] build their probability of the defendant’s innocence in the case of Puckett [22] on the assumption that *‘offenders are distributed equally throughout the population, in terms of race and geography’* – an assumption which has no basis in reality. David Balding dismisses the idea that there is such a thing as a ‘random match’ in the population, by discussing that there are shared genetic traits between relatives, which become weaker as the generational gap grows.

This dependency between supposed unrelated individuals supports the argument that genetic traits are not ‘evenly spaced’ throughout the sample space, but are clumped together until migrated. Redmayne [18] agrees: *‘It is known that allele frequencies vary between races, and for this reason forensic science agencies keep separate databases for each of the major races, for example Afro-Caribbean, Asian and Caucasian. It has been argued, however, that each of these racial groups contains several genetically distinct sub-groups. A frequently used example is the United States Hispanic population, which contains people of Mexican, Puerto-Rican, Guatemalan, Cuban, Native American and Spanish descent, with proportions of these varying from region to region. The genetic differentiation of sub-populations is accentuated by the fact that people do not mate at random--choice of mate tends to be conditioned by racial, linguistic, religious and geographical considerations--so allelic*

variation will be concentrated within sub-populations rather than dispersed throughout the population as a whole.’ Since separate databases are kept for each of the major races, and these races intermingle over time, the problem only becomes more and more acute. Buckleton et al acknowledge that the migration issue is becoming more significant in the modern world as humans find it easier to travel around the planet. This fact seems to have been largely overlooked by the Nuffield Foundation: ‘Since DNA is inherited from common ancestors, it is expected that the frequency of profiles will vary across ethnic groups. Some nationalities are more ethnically homogenous than others, but even in ethnically diverse countries like the UK many people will share common ancestors somewhere along the ancestral line. Moreover, people tend to intermarry within smaller groups for geographical, religious and cultural reasons. So, two people within an ethnic group are more likely to have a similar genotype than two people from different ethnic groups.’

Redmayne [18] doubts whether homogeneity can be assumed even within a single ethnic group: *‘In United States v. Two Bulls 925 F.2d 1127 (1991), a Native American database was used to calculate a match probability of 1 in 177,000. As Kaye (Kaye, “Comment: Uncertainty in DNA Profile Evidence” (1991) 6 Statistical Science 196 at p.197) points out, the assumption that the Native American population is homogeneous enough to justify a single general database is dubious’* [18] Still, the Nuffield Foundation [19] are resolute in their belief that homogeneity is a rational assumption: *‘Intuitively, if Defendant D and unknown perpetrator U are from the same population, the probability of a shared genotype is somewhat greater because the incidence of that particular genetic characteristic may be greater in D and U’s shared ethnic sub-population than in the general population. In these circumstances, the probative value of a matching profile will be correspondingly reduced. In mathematical terms, the likelihood ratio would be smaller because the genotype probabilities, forming the denominator of the likelihood ratio, would be larger.’* Yes, the incidence of a shared characteristic ‘may be greater in a sub-population than in the general population’, but the question is, how can this be effectively measured. While ‘intuitively’ the fact may be so, Tribe’s claim of a ‘soft variable’ within this area may hold true.

6.10.2 Is homogeneity even relevant?

Secondly, even in the event that homogeneity can be considered a rational assumption, one must discover whether homogeneity is even in play. What does homogeneity mean? If we take a prior of $1/WP$, this means that ‘every person in the world population has an equal probability of committing the crime’. Does this not also include opportunity? If we consider that the world population is the ultimate reference class - or location evidence – much in the same way that Fenton and Neil’s $1/100$ prior means that there are 100 suspects gathered around the victim, is it not true that $1/WP$ also means the whole world are ‘gathered around the victim’ too? If this is the case, it matters not that there is homogeneity, because every single person in the world population is equidistant away from the victim – in other words, all have equal opportunity as well as equal probability of having committed the crime. They are all the same distance away from the victim as each other!

Therefore, while it may be said that the ‘old woman from Beijing’ may have less opportunity of having committed the crime than a local man, we can only say this separately from the DNA evidence – in effect the DNA evidence and the location evidence cannot and must not be used to pull each other up by their respective bootstraps. As an example, we could use a simple case:

Let us say that in a bar there are fifty people, plus the victim of a bar brawl. There is therefore a reference class of suspects of fifty all who have the same probability and opportunity of committing the crime. Let us now say that a witness states that the defendant was the furthest person away from the victim. It would seem that the defendant had less opportunity of having committed the crime, since he was furthest away, but actually what the witness is doing is creating a separate reference class, containing only the defendant. Which is the correct reference class, 1 or 50? The two pieces of evidence actually contradict each other, since if the witness is deemed to be correct, the larger reference class is irrelevant. While the 50-person reference class evidence seems to be important, it actually is not – the probability and the opportunity are conjoined, yet we assume them to be different concepts.

In effect, this means that homogeneity is irrelevant – it matters not whether or not suspects are ‘perfectly evenly spaced’ around the globe, because in DNA matching, the expert’s prior starts on the premise that every member of the world population is probabilistically and opportunistically equal in being the suspect.

For this reason, any reference class prior - including $1/WP$ - is flawed, unless it is used in the unique manner prescribed in this chapter. Although Aitken and Taroni [38] have recommended the use of reference classes as priors, and have even stated the equality of suspects within that reference class, they have failed to recognise the importance of communicating K to the jury: *'The relevant population is determined from the circumstances of the crime, and refers to the class of individuals to which the criminal, as yet unknown, can be said to belong. This population may be used to help determine--or, more strictly speaking, estimate --the probability of particular evidence, for example DNA frequencies... The probability that a person chosen at random from the relevant population is guilty can be calculated, in the absence of any other information, by dividing 1 by the number of people in the relevant population. Thus, if all we know is that there are 1,000 individuals in the relevant population, of whom one and only one is guilty, the probability of any individual chosen at random being the guilty individual is 1/1000. This may be taken as a numerical representation of the belief that the person chosen at random is just as likely, and no more likely, to be guilty as anyone else similarly chosen at random from the relevant population.'* The problem here for the jury, is how many suspects are there? From the example given here, it cannot be said.

6.10.3 Priors and likelihood ratios for reference class evidence do not seem to exist in some cases:

It has been argued [47] that BT is an appropriate means of gauging the probative value of evidence, regardless of its type or basis. While it may be argued that the LR should be used as a means of presenting the probative value of evidence, there are some cases where it is difficult to ascertain what the LR actually is. Let us take the following case example:

Hp: 'That the defendant is guilty'

Hd: 'That the defendant is not guilty'

E: The defendant was seen at the scene of the crime with three other people

Intuitively, we must calculate the posterior: $P(Hp|E)/P(Hd|E) = 0.25$ - which would equate to *'the probability of the defendant being guilty, given the evidence, is 0.25'* because the defendant has the same probability of being the guilty person as the other three suspects, who all have an equal probability of guilt of 0.25. In this way the posterior is mutually

exclusive/exhaustive. These four suspects are called a ‘reference class’, since they have been ‘ring-fenced’ from the rest of the world population, and each carry an equal probability of guilt. The question is, what is the prior for this evidence, and having found the prior, what is the LR?

Conversely, as an alternative, let us say that the prior $P(H_p)/P(H_d)$ is 50-50 (or 1). This is a good place to start, because as we discussed in Chapter 3, Buckleton suggested that a 50-50 prior might lead to a ‘safe’ prosecutor’s fallacy, and this following example shows exactly why 50-50 priors are almost impossible to justify in any case. BT is $P(H_p|E)/P(H_d|E) = P(H_p)/P(H_d) \times P(E|H_p)/P(E|H_d)$ and from this we can extrapolate the LR as:

$$P(E|H_p)/P(E|H_d) = 0.25$$

In the odds form of BT, this means that the evidence supports the defence hypothesis by 3:1, which should not as the evidence MUST favour the prosecution case. In effect, the evidence cannot be independent of the prior, because whatever the LR is, the posterior is always 0.25.

However, let us say instead that the prior is $1/WP$ (the world population of approximately 7 billion people) instead. This means that the $LR = WP/4$ (hugely in favour of the prosecution hypothesis) because:

$$1/WP \times WP/4 = 0.25$$

But what does $WP/4$ as a LR actually mean? It means that the four suspects were filtered from a group of 7 billion suspects. But what was this process? To put $WP/4$ into perspective, this conditional probability would normally be achieved by finding a common trait between four suspects which would not proliferate in the rest of the world population. Let us say that this trait was ‘DNA with a RMP of $1/0.25WP$ ’ – which would mean that identical traits would be found in around 1 in 1.75 billion people randomly scattered around the world population. $P(E|H_p) - 1$ (assuming a 0 error probability rate) / $P(E|H_d) - 1/0.25WP = WP/4$. Of course, this is just one solution of how to achieve a 0.25 posterior from an independent prior, but as can be seen from the case example, this was not the case when the four suspects were seen at the scene of the crime. These four, including the defendant, share no common

traits which are might be prevalent in the rest of the world population, except for the fact that they were seen together at the scene of the crime.

Therefore, the only conclusion to be drawn from the evidence, in this case, is that there can be no real or meaningful LR in existence for reference class evidence. Of course, this problem also applies to CCTV evidence placing the defendant at the scene of the crime, or databases placing suspects within a geographical location etc.

What is even more of a concern is that since the world population might also be considered a 'reference class', since each member carries an equal probability of guilt, it can also be said that no meaningful LR can be formulated for this evidence either. This discussion has been expanded upon within Chapter 3.

Even if it could be argued that E^2 has an in-built prior of 50/50 instead, and not $1/WP$, this causes a further problem: What is the LR for this piece of evidence? Since the posterior is $1/100$, we can extrapolate from a prior of 1, that $P(E|Hp^2)/P(E|Hd^2) = 1/100$, an increase towards the defence hypothesis, despite the defendant being found at the scene of the crime! It seems paradoxical that the LR should favour the defence hypothesis when it is actually evidence for the prosecution, until E^1 and E^2 are examined together:

Firstly, E^1 : This posterior = $1/7,000$, meaning there are 7,000 suspects with an equal probability of supporting Hp^1

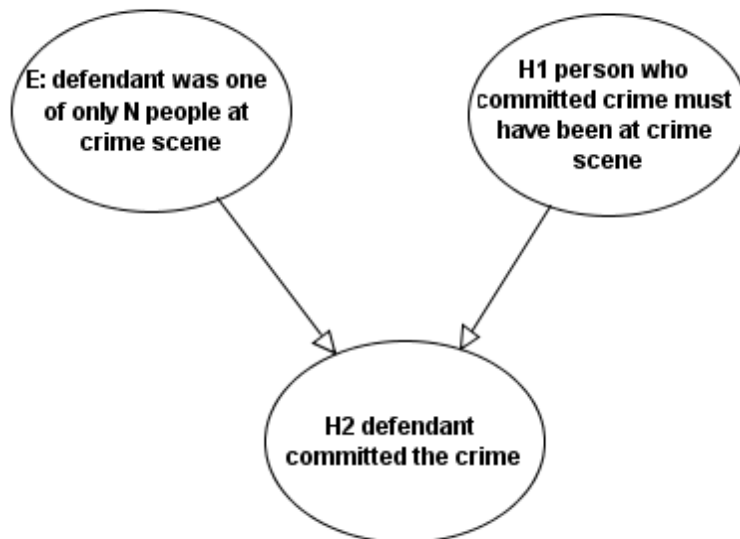
Secondly, E^2 : This posterior = $1/100$, meaning there are 100 suspects with an equal probability of supporting Hp^2

Under circumstances where K is consistent throughout the model, E^2 would be a sub-class of E^1 and would serve to *reduce* the number of suspects to 70. However, where K is not consistent throughout the model, E^2 is independent of E^1 meaning that the number of suspects is *increased* to 7,099 (not 7,100, because the defendant is actually a member of both groups).

Professor Norman Fenton [60], using standard practice, illustrates how cases with evidence such as this are modelled:

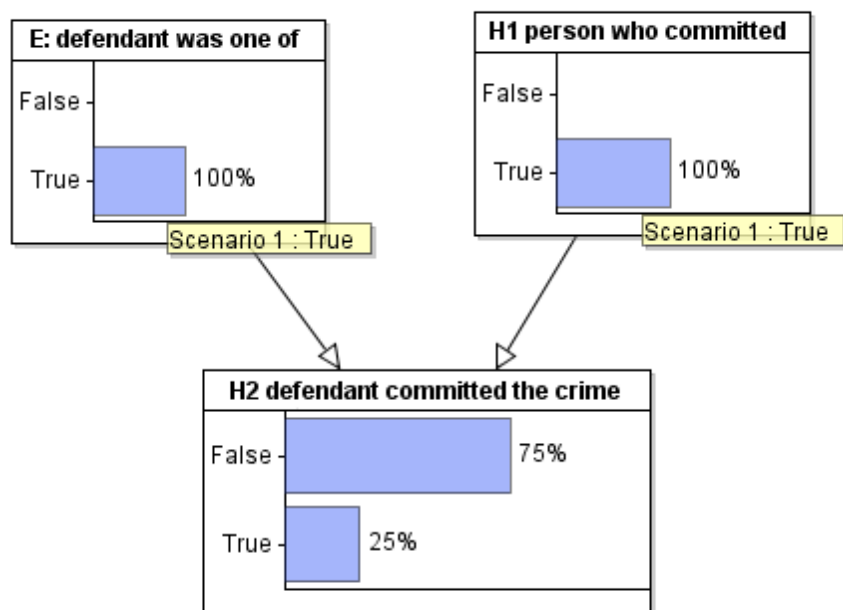
In this simplified model, $H1$ = the probability that whoever committed the crime must have been present at the scene; $H2$ = the probability of 'guilt'; $E1$ = the number of people, of which the defendant is one, observed at the scene who all share an equal probability of guilt.

Fig 6.10.3(a)



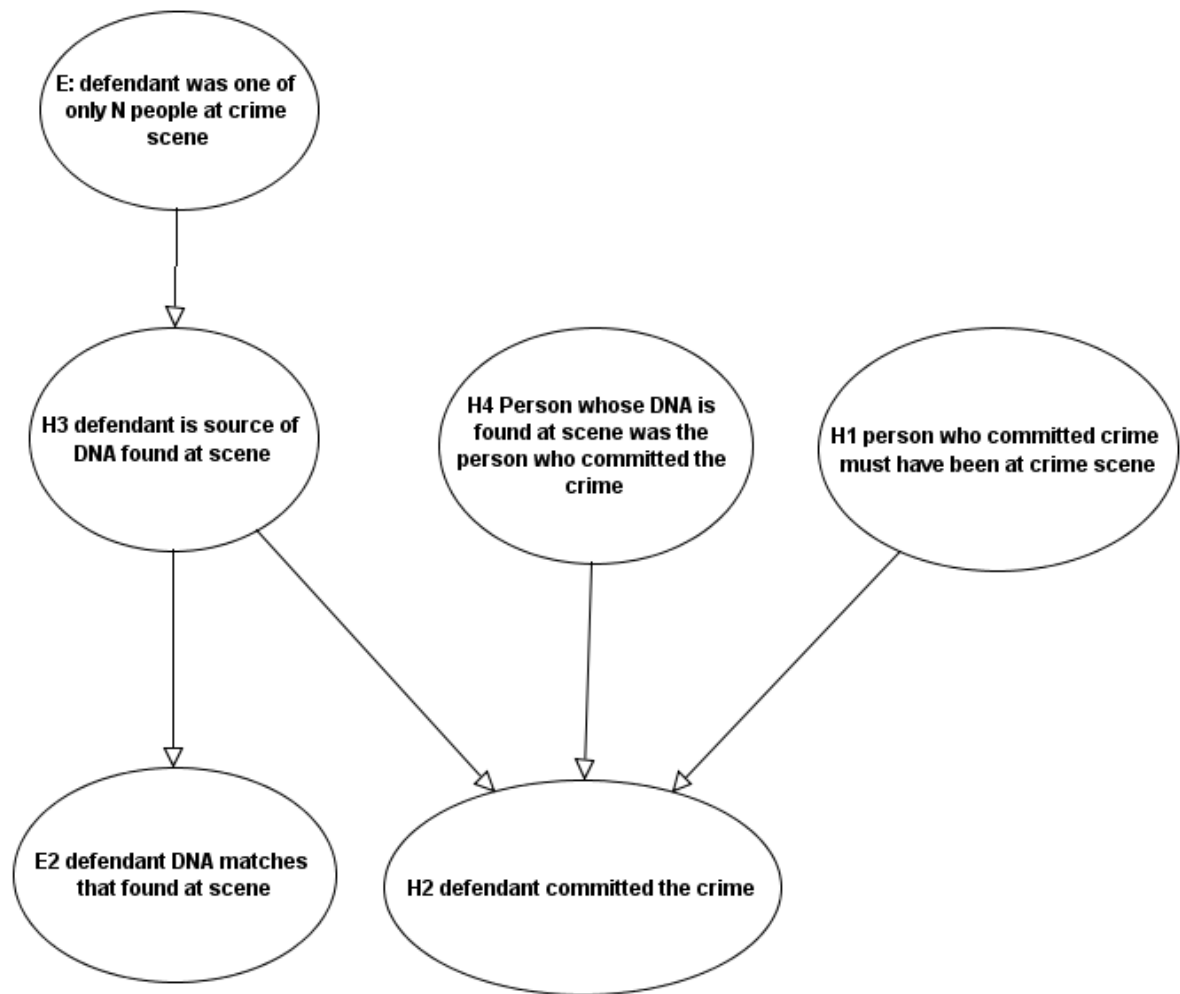
...then after observing E – and assuming $H1$ is true we get:

Fig 6.10.3(b)



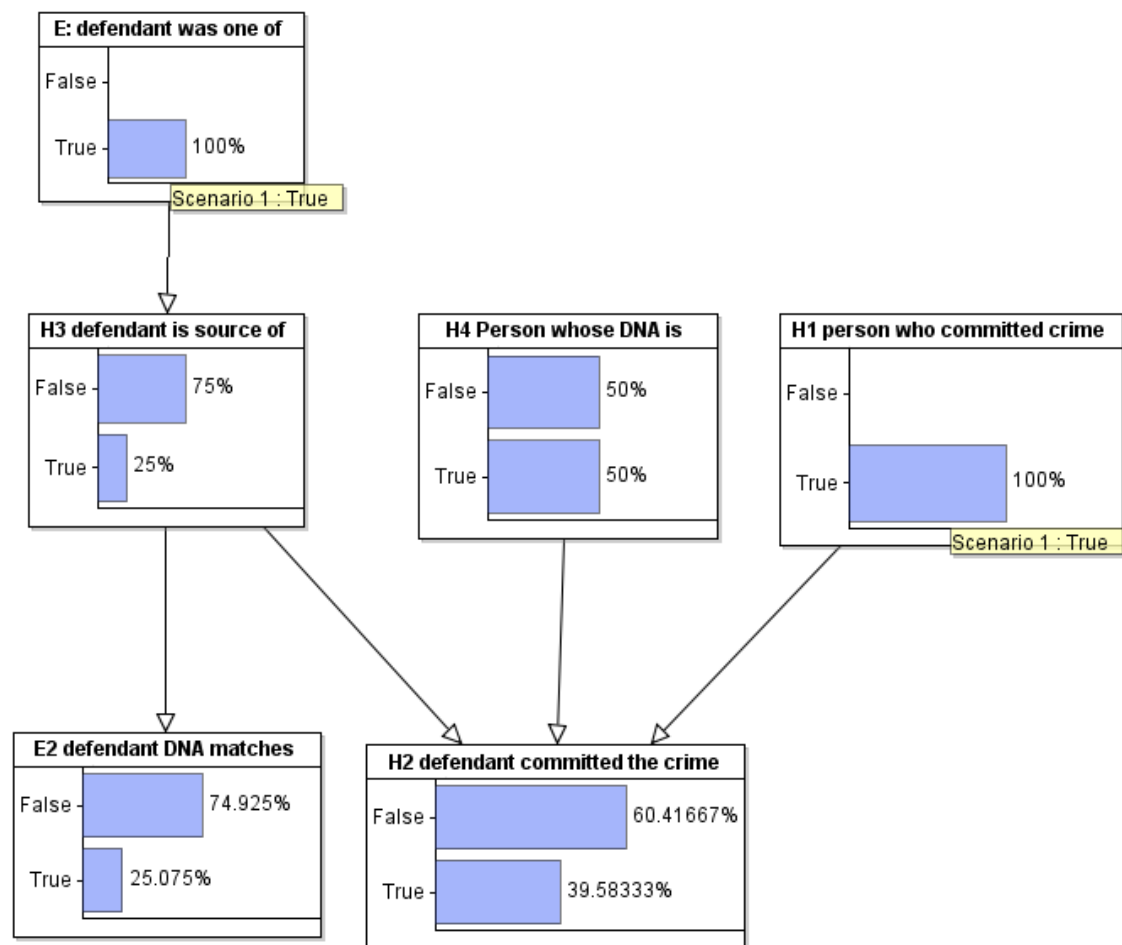
However, the complete model with DNA evidence is this:

Fig 6.10.3(c)



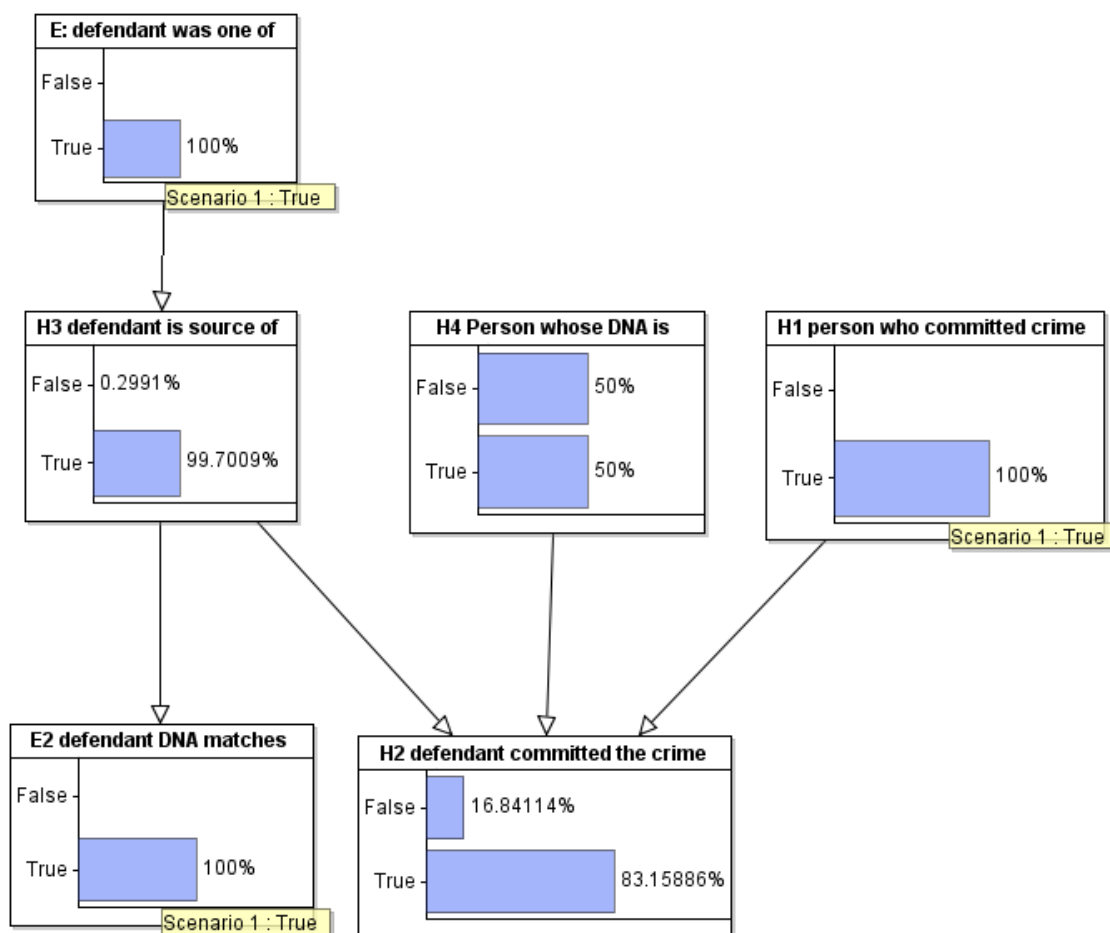
...now if we enter the assumptions and the fact that the defendant was one of four at the scene:

Fig 6.10.3(d)



...and now we enter the DNA match evidence:

Fig 6.10.3(e)



However, as with Figs 6.9 (a)-(d), the problems of incompatible priors and false assumptions of homogeneity and ‘perfectly evenly spaced’ DNA matches still exist. While the posterior of guilt (H2) is at a very high 83.15886% - heavily weighted in favour of the prosecution case – the evidence combination in the model is based on the following assumptions:

1. That H3 is provides a prior of $\frac{1}{4}$, reflecting the evidence that the defendant was one of four people observed at the scene – this is to be taken as fact;
2. That E2 is a DNA RMP, which assumes that in order to combine with H3, the DNA matches are homogenous in the world population and perfectly evenly spaced from each other.

Incredibly, these assumptions also overlook the crucial point that even if we could assume that the DNA matches were homogenous in the world population and perfectly evenly spaced from each other, that the probability of the match closest to the victim not having committed the crime is 0 – a result that no mathematician would allow.

6.11 DNA and the Crown Prosecution Service:

What does this all mean for DNA presented as a RMP as probative prosecution evidence? The CPS [7] has admitted that DNA should not be used alone to convict by suggesting that in a rape case in Huddersfield ‘*an old woman from Beijing*’ sharing DNA profile traits with that of the crime scene trace would be summarily excluded from enquiries. This argument needs to be broken down to examine how much of this evidence can and cannot be combined. Let us say that in this case, there was evidence from the victim that the attacker was a local white male aged 25-30. The evidence in the case can be broken down into:

E^1 : Male attacker

E^2 : Aged 25-30

E^3 : DNA with a RMP of 1/1 million

E^4 : Local to Huddersfield

E^1 , E^2 and E^3 can all be combined, keeping K consistent throughout the model, to form a ‘master RMP’. E^4 cannot be combined with the other pieces of evidence, but is useful evidence to the prosecution hypothesis to keep the defendant ‘live’. The ‘old woman from Beijing’ can be excluded on the basis of not sharing traits E^1 , E^2 and E^3 with that of the attacker, but this is not ‘combining’ evidence in any true sense, in the same way that an attacker’s traits are combined. The reason is that the old woman from Beijing does not really exist unless she is found, located and excluded from enquiries.

6.12 Conclusion:

The CPS claims that DNA is not enough evidence to merit a conviction alone. Since DNA evidence is merely one type of trace evidence, this must mean that evidence presented as a RMP is not enough to merit a conviction alone. The other types of evidence – location/propensity etc - are important because they provide a full picture of the prosecution case. However, since these types of evidence cannot be easily combined with evidence presented as a RMP, for the reason that the hidden value K is not consistent throughout the model, this raises a question: Are these different types of evidence being used to support each other in practice?

In other words, is the CPS using propensity evidence to prove its case on, say, DNA, or using DNA evidence to prove opportunity? The only way to be sure that these dependencies do not occur is to promote transparency of legal reasoning through the medium of BT. We cannot know for certain that a prosecutor will not tell a jury that a person from Huddersfield is more likely to commit a crime in Huddersfield than a person from Beijing, purely on the strength of finding a DNA sample with a small RMP at the scene of the crime. This may be an accurate proposition, but using DNA evidence in this way to prove a completely unconnected point is not accurate.

If any trace evidence, including DNA, is not enough to convict a defendant alone, and if trace evidence is not able to be combined with other types of evidence, then the question is; what really is the probative value of trace evidence? It must be noted at this stage that this author is not questioning that DNA is not powerful evidence in and of itself, but that he is asserting that DNA match evidence should be applied correctly, and in most cases only to exclude suspects from enquiries – not to provide evidence with probative effect towards the prosecution ultimate hypothesis of ‘guilt’.

6.13 Recommendations:

Since it has been shown here that priors reflect the original fact finder’s (the lay or expert witness) reasoning process when valuing the evidence firsthand, this should put a stop to the debate as to what an ‘appropriate’ prior would be in any given situation and for any piece of evidence. Whatever reasoning process it took to formulate the pre-trial probative value of the evidence, this *must* be communicated to the juror in order that the significance of the evidence is not misinterpreted – and in doing so, the oft-silent, but crucial, value *K* will be preserved.

Priors may well be subjective in nature, but in criminal trials, it is the agreement between the subjectivity of the witness and the subjectivity of the juror which is vital. If this express agreement is lost, the juror is left to his own devices as to interpretation of the case before him. While this may leave his role unfettered, it also means that due to his lack of autonomy in being able to gather evidence himself, it also means that his role is greatly uninformed. A

Bayesian model will provide transparency, and eradicate the risk of potential misunderstandings between counsel for the prosecution and defence, and the jury.

When communicating priors, the jury must completely understand the nature of the prior in order that it understands what the posterior means. In DNA match testing, the key information relates to how many suspects there are in the case. By demonstrating here that the number of suspects in the case cannot be reduced by combining RMPs with other types of evidence, a new way of combining evidence must be sought. If not, there is the danger that evidence can be used by the prosecution to support each other, rather than supporting the prosecution case as a whole.

Since it is generally the case that we are dealing with ever-decreasing probabilities, rather than increasing ones, any fallacious reasoning applied when presenting evidence, or any miscalculation of the probative value of evidence tends to favour the prosecution case. This will always carry the risk of miscarriages of justice against the defendant, which cannot be in keeping with the criminal justice system's aim to provide the defendant with the benefit of the doubt wherever possible.

Therefore, the only logical approach is to treat evidence within 'types' (for example, physical traits etc) as separate from each other, and not combine them across the types. In ensuring that this principle is adhered to, value K is preserved and properly communicated to the jury in order that the probative value of the presented evidence is accurately calculated.

To take the example of a prosecution case where the credibility of the defendant is the key issue - for instance, where the alleged event was unlawful sexual penetration and the defence evidence is that the complainant consented to the sexual activity – the jury may be presented with a number of pieces of evidence, including (non-exhaustively); (i) trace evidence of the defendant's presence at the scene; (ii) character evidence of the defendant; (iii) character evidence of the complainant; (iv) location evidence placing the defendant placing the defendant in a reference class in the proximity of the scene. We know that evidence (i) cannot be combined with evidence (ii), (iii) or (iv) without compromising K . However, how do we compare (ii) and (iii) to establish which of the complainant or the defendant is telling the truth? The way to do this is to examine the pairs of hypotheses which make up each piece of evidence.

So, for (ii) and (iii) alone, we may (depending on how the CPS brought the case) be looking at the case as follows:

H_p^0 : That the defendant's account of events is not credible

H_d^0 : That the defendant's account of events is credible

E^1 : The defendant has adduced character witness evidence in support of his account of events

E^2 : The defendant has adverse propensity evidence in support of the complainant's account of events

What is important to distinguish are the hypothesis pairings for E^1 and E^2 , because without doing so, it may be that a jury could fallaciously decide that the stronger their perception of E^1 , the weaker their natural perception of E^2 would be, based on the two pieces of evidence being mutually exclusive/exhaustive (MEE) of each other. Of course, the correct approach would be to 'unpick' the hypothesis pairings for each piece of evidence to ensure that the jury do not create non-existent dependencies.

So, for E^1 , the hypothesis pairing might be:

H_p^1 : That the defendant is generally truthful

H_d^1 : That the defendant is not generally truthful

However, for E^2 , the hypothesis pairing might be:

H_p^2 : That the defendant has committed similar crimes in the past

H_d^2 : That the defendant has not committed similar crimes in the past

It can be seen here that there is no correlation between H_p^1 and H_d^2 , nor is there any correlation between H_p^2 and H_d^1 , yet it may be perceived that there is, especially if the defence do not make an express point, to the contrary, on the matter (despite the fact that it is not incumbent on the defence to disprove any part of the prosecution case). Therefore, it is crucial that the jury are informed of the hypothesis pairings for each individual piece of evidence, for without which, a weak piece of defence evidence runs the risk of inadvertently strengthening the prosecution case, and vice versa. The only way to ensure consistency in

approach, is by treating all pieces of evidence with a Bayesian approach of its strength in the case as a whole. The discussion of mutual exclusivity/exhaustivity (MEE) will be expanded upon in Chapter 7.

Paradoxically, while it has been argued here that under the current practice BNs cannot assist in combining evidence across different types - due to a discrepancy in the value K across the types - Bayesian inference and reasoning has been useful in exposing many common fallacies and misconceptions about the way evidence is combined in complex cases. While this may seem like scant consolation to those who advocate the use of BNs to calculate overall posterior probabilities of guilt in courtroom proceedings under the standard current practice, there can be no doubt that quantifying uncertainty and charting dependencies in evidence in a casual model, such as a BN, can do nothing but assist a jury in establishing whether separate pieces of prosecution (or defence) evidence is being used to support each other to weigh towards the overall hypotheses of guilt or innocence.

7 CHAPTER VII: Problems and limitations with the ‘Lone Likelihood Ratio’ approach

7.1 Introduction:

The focus in this chapter is the likelihood ratio (LR) and its role in criminal trials. To date, the controversy engendered by the use of the LR – a traditional means of presenting the probative value of a piece of evidence - centres on the argument that the LR may or may not be used without a full Bayesian model (ie where a prior and a posterior probability are not considered). This chapter explains that a LR can never be considered in a vacuum and that in a criminal case the full Bayesian application must be provided to a jury in order that the evidence is prescribed its correct weight.

This chapter is a natural continuation of the discussion of the LR and value ‘*K*’ in Chapter 6, but uses case law decisions in cases such as *R v George* [8] and *Regina v Sally Clark* [9] to explain where the ‘lone LR’ has caused problems which would have been eradicated with the use of a full Bayesian model to expose fallacies, errors and misunderstandings related to the evaluation and presentation of the evidence in the case. As a natural consequence of this discussion, the issue of mutual exclusivity/exhaustivity (MEE) is explained, discussed and resolved by an application of Bayesian reasoning.

This author has co-authored and published two papers on this subject [10] [62], and the subject matter in this chapter is not only a key support for the research hypothesis, but is a vital recommendation towards the widespread introduction of BT to the criminal trial process.

7.2 The likelihood ratio (LR):

The ‘odds’ form of Bayes’ theorem (BT), $P(H_p|E) / P(H_d|E) = P(E|H_p) / P(E|H_d) \times P(H_p) / P(H_d)$, where H_p is the prosecution hypothesis, H_d is the defence hypothesis and E is the evidence in the case, shows clearly that the posterior odds of guilt are the likelihood ratio (LR) $P(E|H_p) / P(E|H_d)$ multiplied by the prior odds of guilt. However, as was discussed in Chapter 6, we also know that the (normally silent) value *K* must be consistent throughout the

model. With this in mind, we must also know that for BT to provide accurate probabilistic calculations, the LR is related to the prior.

It has been argued that the LR alone can be used to provide a measure of the probative value (see paragraph 2.7) of a piece of evidence - without using a full Bayesian interpretation - but this argument disregards the fact that the LR was based on a pair of hypotheses, which therefore forms its natural prior. For example, a DNA sample presented with a random match probability (RMP) of 1/1 million, has a LR of 1 million (because $P(E|H_p) = 1$ (assuming a 0 error probability rate), and $P(E|H_d) = 1/1\text{million}$). What this means is that the DNA evidence, in the absence of any other evidence in the case, supports the prosecution hypothesis by a factor of 1 million. This seems very probative towards the prosecution hypothesis – and it is – until the full value of the evidence is presented with its natural prior of 1/WP based on the following hypothesis pairing:

H_p: ‘The defendant was the source of E’

H_d: ‘Someone else in the world population, alive at the time of the crime, was the source of E’

E: DNA with a RMP of 1/1 million

Once the prior is factored-in, the probative value of the evidence is put into a better perspective: There are 7,000 people in the world population with the same evidential traits, which means that *with no other evidence in the case*, this evidence actually supports the defence hypothesis by a factor of 7,000. In theory, the LR of this evidence can be used without express consideration of the prior, but for the jury to understand what the LR of 1 million actually means or pertains to, the value *K* must be communicated to them by the expert - which involves a discussion of the prior. Therefore, a full Bayesian interpretation of evidence cannot be avoided in any case.

Conversely, let us say that the LR of a piece of DNA evidence is presented as 1/1 million, but *K* is not communicated to the jury: $P(E|H_d) = 1/1\text{million}$, which means that ‘*the probability of seeing the evidence given that the defence hypothesis is true is one-in-a-million*’. Without the communication of *K* (and therefore the prior of 1/WP), 1/1 million could wrongly be taken to mean ‘one in a million suspects’, instead of the accurate ‘one in *every* million suspects’ – a 7,000-fold increase in the suspect pool. At one point, it was thought that the LR

could be considered without the prior. Fenton and Neil discuss the advantages and disadvantages of using the LR without adopting the full Bayesian calculation: *'Another advantage of using the likelihood ratio, is that it removes one of the most commonly cited objections to Bayes Theorem, namely the obligation to consider a prior probability for a hypothesis like 'guilty'. Hence, the use of the likelihood ratio goes a long way toward allaying the natural concerns of lawyers who might otherwise instinctively reject a Bayesian argument on the grounds that it is intolerable to assume prior probabilities of guilt or innocence.'* While this statement is true, because a hypothesis of 'guilt' may not have a direct correlation to the hypothesis pairing used by the witness, it is nonetheless untrue to say that a piece of evidence can be considered away from its OWN hypothesis pairing.

For the LR of a piece of evidence to provide an accurate probative value, the hypothesis pairing it refers to MUST be a true negation of each other [10], which means that the pairing must be mutually exclusive and exhaustive of each other. What this means is that the hypothesis pairing must show an exactly opposite view of the case, which must also encompass every possible eventuality. If it does not, the evidence cannot provide an accurate probative value towards *either* hypothesis.

The Nuffield Foundation [16] recognises the importance of mutual exclusivity/exhaustivity as follows: *'Probabilistic calculations of (un)certainly obey the axiomatic laws of probability, the most simple of which is that the full range of probabilities relating to a particular universe of events, etc. must add up to one.... Where there are two exhaustive and mutually exclusive possibilities, the probability of one can always be calculated if the other is known, e.g. $p(\text{Guilty}) = 1 - p(\text{Innocent})$; and vice versa, $p(\text{Innocent}) = 1 - p(\text{Guilty})$...In relation to fact-finding in criminal proceedings, this will often be G, guilt; or I, innocence. Since it is certain that the accused is either factually guilty or factually innocent (there is no third option), $p(G|E) + p(I|E) = 1$ (meaning that the probability of Guilt, given the Evidence; plus the probability of Innocence, given the Evidence, logically exhausts the range of all eligible possibilities).'* It is this acceptance that all events in the universe 'must add up to one' which supports the assertion that LRs must be mutually exclusive/exhaustive if they are to provide an accurate probative value towards a given hypothesis. Confusingly, the Foundation [30] later asserts that while mutual exclusivity is vital, exhaustivity is not: *'What an expert should be in a position to offer is the assessment of a likelihood ratio (LR) for the evidence. The LR is the ratio of two probabilities, conditioned on mutually exclusive (but not necessarily*

exhaustive) propositions.’ However, this must be a mistake as mutual exclusivity naturally assumes that all possible outcomes have been considered and are accounted for. Any other result means that all events in the relevant universe will not ‘add up to one’, which leaves room for error and uncertainty.

When dealing exclusively in LR’s, rather than using a full Bayesian interpretation, it is easy to forget that the evidence pertains directly to a hypothesis pairing. For example, with DNA sampling where the identity of the sample donor is unknown, the hypothesis pairing used to formulate a RMP is:

Hp: ‘The defendant is the source of E’

Hd: ‘Someone else in the world population, alive at the time of the crime, was the source of E’

E: DNA with a RMP of 1/1 million

We know this to be true, because the pairing is mutually exclusive/exhaustive – if the defendant did not leave the sample, then someone else in the world population (alive at the time of the crime) had to have done. This issue of mutual exclusivity/exhaustivity has been recognised by Ward Edwards [12] when evaluating the evidence in *People v Collins* [5] when he states: ‘*Some probabilities are independent; they are conditional only on the model in which they appear and not on any of the specific entities (eg hypotheses) within that model. Two events are conditionally independent of each other if knowing one does change the probability of the other, but that dependence would not exist if you knew the true state of some third event that links them. For example, in Collins, the way the female wore her hair is properly modelled as conditionally independent of whether or not the male driver was bearded, and vice versa. What links these two observations is the possibility that the perpetrator was Janet Collins. Evidence that the mugger had a blonde ponytail, as Janet Collins did, makes it more likely that the driver was Malcolm Collins, who was bearded. If you knew either that Janet Collins was or that she was not the mugger, being told either of these two observations would not change the probability of the other.*’ In effect, mutual exclusivity/exhaustivity ensures that multiple pieces of evidence in a case only appear once - and crucially, are mistakenly taken to be independent of each other when they are not.

The prior of $1/WP$ denotes the prior probability before evidence is presented, which in the current population means a pool of suspects of approximately seven billion people - Buckleton et al [6] support this view. What this means is that the LR for any piece of DNA evidence presented as a RMP comes with an in-built prior which was the prior used by the sampling and testing experts arriving at locus probabilities, and it is the prior which will be adopted by the juror when the evidence is presented in court. The Foundation [30] confirms the position: *‘In a forensic context, the LR can be explained generically as the ratio of: (1) the probability of the forensic scientist’s observations, if the postulated fact in issue were true; to (2) the probability of the same observations if the fact in issue were false’*. In no way can it be said that the LR of this evidence is used ‘without a full Bayesian interpretation’, because the Bayesian interpretation has already been considered.

In effect, the prior $1/WP$ is itself a RMP, which when combined with a LR, provides a posterior ratio which reflects the number of potential matches in the new sub-class – still a RMP, albeit of the number of suspects in the smaller pool.

We can also ensure that $1/WP$ is the correct prior for RMP evidence by considering an alternative hypothesis pairing of:

Hp: ‘The defendant is guilty’

Hd: ‘The defendant is not guilty’

E: DNA with a RMP of $1/1$ million

Let us say that there is evidence that the defendant was seen physically struggling with another man and the victim. In this case $P(E|Hd) = 1$, and not $1/1$ million, because the *probability of seeing the evidence if the defendant is innocent is 1*. In other words, the LR of the DNA would be 1 and not 1 million. This would, of course, mean that the DNA evidence is ‘neutral’, and would likely be excluded from the jury. In Bayesian terms, any evidence which leaves the prior unchanged is neutral.

Neutral evidence is summarily excluded from the jury in the interests of justice to either expedite trial proceedings or to reduce the risk of unfair prejudice from the jury inaccurately valuing the evidence due to routine and institutional counter-intuition. In this case example, since the evidence would be found in every member of the pool (in this case two) of suspects,

it helps neither the prosecution nor defence hypotheses. It therefore follows that Bayesian inferences drawn from RMP evidence based on a $1/WP$ prior cannot be used on sub-sections of the population.

The concept of the apparent ‘neutrality’ of trace evidence was debated in the case of *R v George*:

7.3 *R v George*:

In the case of *R v Barry George* [8] the evidence was a single particle of firearm discharge residue (FDR) found in the defendant’s coat pocket a year after the gunshot murder of television personality Jill Dando. At trial, the FDR was presented as having a $1/100$ probability of being found in the defendant’s pocket if he was innocent, and he was found guilty by the jury. After the trial, Dr Ian Evett, who from 1996 to 2002 worked for the Forensic Science Service (‘FSS’) and who had been developing an evidence analysis technique called Case Assessment and Interpretation (‘CAI’) - which adopts BT as its mathematical analysis tool - invited Mr Keeley, the prosecution expert, to apply the CAI technique to the FDR evidence on two different propositions:

Hp: ‘That the defendant was the man who shot Jill Dando’;

Hd: ‘That the defendant was not the man who shot Jill Dando’

E: A single particle of FDR found in the defendant’s coat pocket a year after the alleged crime

Mr Keeley estimated that the likelihood of his finding one – as had happened in this case - or a few particles, as 1 in 100 *on either proposition*.

$$P(E|Hp) = 1/100$$

$$P(E|Hd) = 1/100$$

$$P(E|Hp)/P(E|Hd) = 1$$

The significance of this was that, in Mr Keeley's opinion, the finding that he made of a single particle had in fact been 'neutral', which he then confirmed to Dr Evett. On the strength of this finding, the defendant was retried without this 'neutral' FDR evidence, and acquitted.

In this case $P(E|H_p)/P(E|H_d) = 1$, because the probability of seeing the evidence under either proposition is equal, which means that the evidence is neutral. However, it is argued [10] that since $P(H_p)$ and $P(H_d)$ in this case are not mutually exclusive/exhaustive, that the evidence is not neutral when looking at the case as a whole. The reason is that the evidence could support both $P(H_p)$ and $P(H_d)$ if, say, the defendant was not the man who shot Jill Dando yet *still* was found to have the evidence in his coat pocket. In this way the LR of the evidence provides inaccurate values and may unfairly weigh the evidence towards one proposition or another.

The only way to an accurate probative value for this evidence, and therefore mutual exclusivity/exhaustivity, is to ensure that the background propositions which lead to the LR are true negations of each other and cover every possible eventuality. To do this, one must examine what the background propositions actually are. Evett et al [63] contrast 'source-level propositions' (SLP) with 'offence-level propositions' (OLP). The Nuffield Foundation [16] even break the levels into four: *'A useful starting point in evaluating expert evidence is to identify the level of proposition (or type of answer) which the evidence addresses. Four different levels of proposition can usefully be distinguished: (i) source level propositions; (ii) sub-source level propositions; (iii) activity level propositions; and (iv) offence level propositions.'* The above hypothesis pairing is obviously an OLP because it compares the defence hypothesis of 'guilt' with the prosecution hypothesis of 'innocence'.

The key difference between this type of pairing and a SLP pairing is that with a SLP the pairing disregards the defendant's innocence in favour of calculating the probability of the defendant being the original source of the evidence. 'Guilt' and 'source' are obviously not mutually exclusive/exhaustive states.

In DNA testing where the evidence is presented as a RMP, this would be based on SLP, since the defendant's guilt is disregarded in favour of calculating the probability of whether he was at the scene of the crime and left the crime scene trace – his guilt may be inferred from his presence, but cannot be ascertained simply by examining the DNA evidence alone.

Therefore, presenting DNA evidence on an OLP pairing is misleading because it implies that guilt can be assumed if the juror is satisfied that the evidence pertains to the defendant.

In George, the FDR evidence is obviously trace evidence, because it is probative towards the prosecution hypothesis of ‘presence’. The key difference in probabilistic terms between OLP and SLP pairings is that for an OLP pairing, the prior is 50/50 (or 1) based on the two states of ‘guilt’ and ‘innocence’; whereas for a SLP pairing, the prior is $1/WP$ based on the seven billion states of ‘defendant is the source’ and ‘someone else in the world population alive at the time of the crime is the source’. This distinction is not easy to make.

Redmayne [18] plainly has also shown the confusion between ‘match evidence’ such as DNA, and ‘guilt evidence’: *‘Discrepancies between two profiles which are said to match can considerably decrease the strength of DNA evidence. On a Bayesian analysis, the correct approach for an expert testifying in a criminal case is to provide the jury with a likelihood ratio which represents: ‘the probability of the evidence given guilt’/‘the probability of the evidence given innocence’.* From the appeal report in George [8], the appeal court interpreted the FDR to be neutral on the following proposition pairing:

Hp: ‘The FDR came from the gun that killed Jill Dando’

Hd: ‘The FDR came from some extraneous source’

Of importance to note from the appeal court’s hypothesis pairing is that it arguably interpreted the evidence as a SLP pairing, not the OLP pairing that Keeley based his LR on. If this is true, then the likelihood of a juror, unused to dealing with forensic evidence has a higher risk of making a similar error than an experienced appeal judge. On a SLP pairing, $P(E|Hp) = 1/100$, which means that *‘the probability of seeing the evidence, given that the defendant is the source is $1/100$ ’*. This makes much more sense than a OLP pairing $P(E|Hp) = 1/100$, which means that *‘the probability of seeing the evidence, given that the defendant is guilty is $1/100$ ’*. With this in mind, $P(E|Hp)/P(E|Hd) = 1$, which means that the evidence does not discriminate between any member of the world population, as the probability of finding it would have been the same on whoever would have been the defendant.

In this way, there is no material difference between the LR of this evidence and a RMP of 1 for any type of trace evidence, including DNA. This seems intuitively incorrect until the hypothesis pairing for trace evidence, presented as a RMP, is re-examined:

Hp: 'The defendant is the source of the evidence'

Hd: 'Someone else in the world population, alive at the time of the crime, is the source of the evidence'

The LR for this pairing is $P(E|Hp)/P(E|Hd) = 1$, which means the probability of seeing the evidence is the same whichever hypothesis is true. In other words, the same evidence can be found in every member of the world population, meaning *that the evidence does not discriminate between any member of the world population*. In George's pre-appeal report prepared by the FSS, it was stated that due to a large number of modern factors, including lifestyle issues and investigative practices, the probability of finding single particles of FDR in the atmosphere is high, and that the probability of these particles being attributed to the person it was found on is low.

Therefore, it is easy to see how the court could have made the finding that the probability of finding the same piece of evidence on any defendant being investigated would be the same – thereby making the evidence 'neutral'.

There were many other non-mutually exclusive/exhaustive hypothesis pairings mentioned by the prosecution throughout the case:

1. Para 18:

Hp: FDR came from gun that killed victim

Hd: FDR came from some extraneous source

2. Para 22:

Hp: FDR recovered from BG's coat pocket, as he was the killer of the victim

Hd: FDR recovered from BG's coat pocket, but he was not the killer of the victim

3. Para 23:

Hp: BG is the man who shot victim

Hd: BG had nothing to do with the incident

4. Para 26 (i):

Hp: FDR came from a gun fired by BG

Hd: FDR came from some other source

5. Para 26 (ii):

Hp: The wearer of the coat fired the gun

Hd: The wearer of the coat did not fire the gun (the defence hypothesis is unspecified in the appeal report, but this is one possibility.)

6. Para 27:

Hp: FDR came from a gun fired at the time of the victim's murder

Hd: FDR came from some other source

7. Para 28:

Hp: FDR found as a result of BG firing a gun

Hd: FDR found as a result of secondary contamination

8. Para 32:

Hp: The particle is FDR

Hd: The particle is not FDR

9. Para 33:

Hp: FDR came from ammunition that killed the victim

Hd: FDR came from any other ammunition that had that kind of percussion primer

10. Para 37:

Hp: That the FDR did not come from secondary contamination

Hd: That the FDR came from secondary contamination

11. Para 38:

Hp: FDR came from the cartridge that killed the victim

Hd: FDR came from some innocent source

12. Para 50:

Hp: That the FDR was deposited on the coat other than innocently

Hd: That FDR was deposited on the coat innocently

Although these pairings did not result in express probabilistic calculations, is it easy to see how a jury might become confused as to what the content of the prosecution case actually is, and start to ‘accumulate’ pieces of prosecution evidence, which, while not independently in support of Hp, might, and are likely to be, misinterpreted to support Hp when combined. For this reason it is vital that a formal probabilistic model is used as an aid to ensure that evidence is given its correct weight and is not counted multiple times.

Absence of evidence:

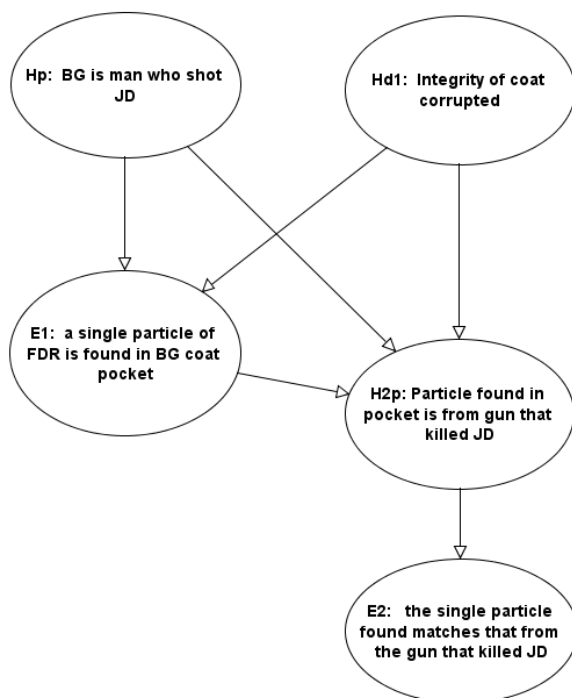
Of course, the FDR in George could only be said to be neutral if it is not probative towards either the prosecution or defence hypotheses. Even if the evidence did not discriminate between members of the world population, does this still mean that the evidence is neutral? Since the prior for SLP evidence is $1/WP$ and the LR is 1, thereby formulating a posterior of $1/WP$, does this not actually support the defence hypothesis? Buckleton et al [6] identify the principle of ‘absence of evidence’ (to be contrasted with ‘evidence of absence’) which refers to a situation where some evidence has been searched for but not found. Buckleton uses to BT to interpret this evidence:

$$LR = P(\text{noE}|H_p)/P(\text{noE}|H_d)$$

The Bayesian interpretation is that ‘...*unless some very special circumstances pertain, then the finding of noE will be more probable under Hd, and hence the absence of evidence supports Hd. Often in real casework, this is only weak support for the Hd.*’ However, in light of the exclusion of ‘neutral’ evidence in cases such as George, the fact that the jury will never get to hear that the defendant was searched and found to have no more evidence than would be expected to be found on any other member of the world population, is something which carries the risk of unfairly weighing the case as a whole in favour of the prosecution hypothesis.

In other words, if the jury only get to observe SLP evidence, then this evidence will certainly support the prosecution hypothesis (there can conceivably be no SLP evidence placing the defendant at the scene of the crime which is actually exculpatory). It would be very probative towards the defence hypothesis in George, and similar cases, that the even after a rigorous forensic investigation, that only a single particle of evidence which does not discriminate between any member of the world population was found on the defendant – certainly more than ‘weak support for the Hd’. In their paper “*When ‘neutral’ evidence still has probative value: implications from the Barry George Case*”, Fenton, Berger et al [10] recommend the following causal BN model to show the dependencies between the hypotheses and evidence in the George case:

Fig 7.3



One of the dangers of using LR_s which are not actually mutually exclusive/exhaustive is that evidence can be said to favour a hypothesis over its negation, without actually supporting the ultimate prosecution hypothesis of guilt. The risk is that a jury might believe that since the LR is mutually exclusive/exhaustive, that evidence with probative value towards Hp means

that the jury must convict. This may have been the case in cases such as *R v Sally Clark*, *The Popi M* and *Nulty v Milton Keynes*.

7.4 R v Sally Clark:

In the case of Clark [9] a mother was convicted for the murder of her two infant sons born a year apart. The prosecution raised the point that both babies could not have both died of sudden infant death syndrome (SIDS) - which the prosecution expert witness, Dr Meadows, sought to prove on statistical grounds. In his evidence, Dr Meadows stated that a single SIDS death occurred in 1 in 8,500 live births in the UK, which meant that, using the product rule, the probability of there being two SIDS deaths in one family was in the region of 1/73 million.

Unfortunately for the defendant, Dr Meadows committed the prosecutor's fallacy [15], which may have been instrumental in the jury's decision to convict. Since her successful second appeal (on a different point, concerning the wrongful non-disclosure of a test result indicating a possible cause of death), a point relating to some of the fundamental probabilistic flaws in the prosecution case engendered much debate concerning the use of the LR in cases such as this. Philip Dawid [64] argues that the probability of the double SIDS deaths should have been compared with the alternate hypothesis of murder, which occurs in approximately 1 in 45,000 live births in the UK. Therefore the case would be as follows:

Hp: 'That the defendant's two babies were murdered'

Hd: 'That the defendant's two babies died of SIDS'

E: The two dead babies

By BT, Dawid argues that $P(E|Hp)/P(E|Hd) = \text{over } 100:1$ against murder. Norman Fenton [65] argues that the true negation of Hp is: Hd: 'That at least one baby died of SIDS', which provides a much lower LR of 1:2.5 against murder – far more probative towards the prosecution hypothesis than Dawid's formulation.

Dawid and Fenton both acknowledge the same basic (and unlikely) assumption that murder and SIDS are the only possible alternative causes of death. The fundamental problem with

this assumption is that H_p and H_d are not mutually exclusive/exhaustive, which means that any LR favouring one hypothesis over the other may be misinterpreted to mean that the evidence favours the ultimate hypothesis pairing of the defendant's guilt or innocence. The only true negation of 'SIDS deaths' is 'not SIDS deaths', while the only true negation of 'death by murder' is not death by murder'. If we take LRs based on these two pairings, we have:

H_p^1 : 'That the babies did not die of SIDS'

H_d^1 : 'That the babies died of SIDS'

H_p^2 : 'That the babies were murdered'

H_d^2 : 'That the babies were not murdered'

E: The two dead babies

The first problem here is that Dawid compares 'deaths per live births', which cannot provide values related to the actual hypotheses, for which we would need to know the number of SIDS/murder deaths per number of infant deaths. Secondly, since there are dependencies between H_p^1 and H_d^2 , and also H_d^1 and H_p^2 it is easy to see that SIDS and murder cannot easily be compared with each other to provide a meaningful LR. Even if these two states are comparable, the only way to gauge the LR's impact on the case is to combine the LR with the prior. If the LR is based on SIDS/murder, and we know for value to K to be consistent throughout the model, the prior must also be:

H_p^3 : 'That the babies were murdered'

H_d^3 : 'That the babies died of SIDS'

E: The autopsy reports on the defendant's dead babies, together with the other prosecution evidence in the case

The problem with this pairing is that the evidence must be annexed to it, but Dawid states that the evidence is '*that the babies have died*', and does not explain how this evidence is annexed to the hypothesis pairing. In fact, the evidence Dawid uses is statistic related to the number of SIDS deaths/murder deaths per live births in the UK, which is based on other babies that have died - not the defendant's. The Nuffield Foundation [16] also seems to endorse the use of evidence of other cases to support probabilistic values in the (very separate) case under consideration: '*It might be very unlikely that two cases of SIDS would be experienced in a*

single family. But it might be even less likely that a mother would serially murder her two children (we must make assumptions here, of course, about the impact of other evidence). So, taken in isolation, the bare fact of two infant deaths in the same family is probably more likely to be SIDS than murder. Unlikely though the former innocent explanation may be, it is not as unlikely as the latter, incriminating explanation.'

If the evidence favours neither the prosecution or defence case, then $P(E|Hp^3)/P(E|Hd^3) = 1$ - which means that the posterior remains at the same value as the prior. But what is the prior in this scenario? For the legal principle 'innocence until proven guilty' to apply, the posterior must reflect innocence.

Let us say that the ultimate hypothesis pairing in the case should be:

Hp^4 : 'That the defendant is guilty'

Hd^4 : 'That the defendant is not guilty'

E: The autopsy reports on the defendant's dead babies, together with the other prosecution evidence in the case

Hp^1/Hd^1 and Hp^2/Hd^2 do not reconcile with Hp/Hd^4 . The only rational answer is that the 'number of SIDS/murder deaths per live births' is unrelated to the case, and that the defendant should be tried on the strength of the prosecution evidence against her, rather than on statistics relating to other cases that she has no knowledge of, or can defend herself against. This principle has been recognised by James Franklin [42] who argues: *'...probabilities most usually and easily quantified, as those arising from a proportion in a reference class like $P(\text{this } A \text{ is a } B | x\% \text{ of all } As \text{ are } Bs) = x\%$, are of very doubtful legal relevance at all. The evidence that 99% of drivers cut a corner is not allowed as evidence that a particular driver cut that corner on a particular occasion. It is neither allowed as sufficient evidence to prove that hypothesis nor allowed as partial evidence. While there have been debates on the admissibility of certain kinds of other 'similar fact' evidence such as evidence of prior convictions and of character, there has been no serious suggestion that simple membership of a reference class of high criminality (or high criminality of a certain sort) should be admissible as evidence.'*

The same problem issue in Clark can be seen in the case of Nulty [66], where the three alternative causes of a fire were presented as a discarded cigarette end, a faulty electrical cable, or arson. The court in Nulty applied the reasoning from the maritime civil case of the Popi M [67], which based a LR of a sinking ship on the opposing hypotheses of the event being caused either by the negligence of the crew, or a peril of the sea, such as passing submarine.

The judge in the case ruled out any evidence of wear and tear, and so felt obliged on the balance of probabilities to prefer the negligence hypothesis. The reason the judge gave, was that there was no other hypothesis or evidence presented which gave a viable alternative. The problem with this analysis is that neither of these hypotheses may have reached the required standard of proof.

7.5 Conclusion and recommendations:

As can be seen in this chapter, evidence in a complex case has the potential for being evaluated and presented in a way which may be misinterpreted as to its weight and/or relevance to the case as a whole.

The interesting thing to note, is that it is plain and obvious that mistakes are routinely made where prosecution and defence hypotheses are somehow ‘divorced’ from the evidence purported to support them. In order to ensure that the jury is under no misapprehension as to the importance of a single piece of evidence, it must be viewed within the context of a full Bayesian model to show where the inherent dependencies and uncertainties are.

If the full Bayesian model is not used at both the pre-trial evidence evaluation stage, or at the in-trial evidence presentation stage, there is a real danger that evidence may not only be ascribed the wrong weight, as in *Regina v Sally Clark*, but, as in *R v George*, it may be precluded from the jury’s decision-making process entirely.

Of course, these two cases are the main focus of this chapter, but it should not be inferred that this is an insignificant problem with narrow application. It is simply that the problems

discussed in these two highly publicised cases rarely come to light and are almost never reported unless they are identified, exposed and become the subject matter of an appeal.

8 CHAPTER VIII: The Full Code Test

8.1 Introduction:

The focus of this chapter is the recommendation of a Bayesian approach to the pre-trial evaluation stage of evidence by the Crown Prosecution Service (CPS) and at the mid-trial ‘submission of no case to answer’ (SNCA) application by the defendant, in order to ensure that there is a consistent, transparent, formalised framework to the decision to prosecute and/or continue with the case mid-trial.

The chapter begins by explaining the threshold for the bringing of cases, which can be found in the Full Code Test for Crown Prosecutors and explains the flaws in this threshold which leaves wrongly-tried defendants without an effective remedy against the CPS - apart from the unsuitable remedy of defending themselves in a court which they should not have been in in the first place. The key case in this area is *R(B) v DPP*, [68] the ruling of which explains this apparent flaw in the system. There is also an analysis of the SNCA threshold for the same reasons.

This chapter continues the examination of the case of *R v George* [8] – of which the problems of the ‘lone LR’ ascribing the potentially wrong weight of the evidence in that case, were discussed in depth in Chapter 7 - and explains how, after his retrial acquittal, the Full Code Test threshold left that defendant without any effective form of redress under the current system. Of course, this discussion in no way undermines the findings in Chapter 7 about the probative weight of the evidence in that case, or whether that evidence should have been precluded from the jury in the retrial.

This chapter not only forms the key support for the recommendation of Bayes’ theorem for the research hypothesis, but should encourage further debate regarding reforming the Full Code Test with the addition of a Bayesian element.

8.2 The Full Code Test for Crown Prosecutors:

The Full Code Test for Crown Prosecutors (FCT) [69] is issued by the Director of Public Prosecutions (DPP) which aims to regulate the CPS's decision to prosecute, by setting the threshold by which cases should or should not be brought. The test should act as a filter to prevent unmeritorious cases from reaching the trial stage, thereby saving public expense and providing the means by which citizens may seek redress against prosecutors who '*should not start or continue a prosecution which would be regarded by the courts as oppressive or unfair and an abuse of the court's process*', as set out in FCT Code 3.5. The test is applied by each prosecutor individually at any time during the police investigation process, and before trial, and, as has been admitted [70] by the DPP policy director, there is no set formula for a consistent application.

During the trial, the defence have the option of applying to the judge to have the trial stopped for lack of evidence in an application of 'no case to answer'. For a defendant to succeed in this mid-trial application he must persuade the judge that there is insufficient evidence to proceed, and in this way he, again, may seek redress for an abuse of process if his application is wrongly denied. However, as with the FCT, there is no formal framework for an application of this test.

The problem is that since lack of weight of evidence is the key to a defendant avoiding proceedings, and that it is well known [71] that lay people routinely attach inaccurate weight to evidence due to biases and acceptance of fallacious reasoning [13] [5] [8] [1] [14], these tests are not fit for purpose. One suggestion for achieving a consistent and transparent application of these tests is to apply Bayesian probabilistic reasoning to the evidence in the case and allowing the judge, prosecuting and defence counsel to fully and accurately evaluate the case against the defendant in order to make decisions to prosecute, and submissions of no case to answer, more efficient.

8.3 The threshold:

The FCT has two limbs as set out in Code 3.4; firstly, the 'evidential stage', where the evidential threshold for the bringing of a case is considered, and secondly, the policy test which allows cases which has passed the evidential stage to not be permitted to proceed on public policy grounds – for instance where the penalty is too low to justify the expense of

going to trial. The evidential stage sets out the threshold as: *‘Prosecutors must be satisfied that there is sufficient evidence to provide a realistic prospect of conviction against each suspect on each charge... A “realistic prospect of conviction” is an objective test based solely upon the prosecutor’s assessment of the evidence and any information that he or she has about the defence that might be put forward by the suspect. It means that an objective, impartial and reasonable jury or bench of magistrates or judge hearing a case alone, properly directed and acting in accordance with the law, is more likely than not to convict the defendant of the charge alleged.*

In the case of *R(B) v DPP* [68], the High Court ruled that the prosecutor must take a merits-based approach to the decision to prosecute, rather than a ‘bookmakers’ approach, which means that the prosecutor must place himself in the position of a reasonable, impartial, objective – and therefore, hypothetical – juror and ask whether, on balance, the evidence was sufficient to merit a conviction; rather than applying a purely predictive approach based on past experience of similar cases, since this would mean that cases which may contain enough evidence but which are notoriously difficult to prosecute – for instance, where there is no third party witness or external evidence to a crime alleged to have been perpetrated between by one person and another – would never be brought. However, although the prosecutor has a duty to make a probabilistic decision (*‘on balance’*) the test combines an objective assessment (*‘enough evidence’*) with a partially subjective one, where they place themselves *‘in the position of a juror’*, a position which allows any uncertainty in the evidence to be subsumed within the grey area between what the prosecutor may believe about the merits of the case, and what he thinks a hypothetical jury might believe. As a side note, this ruling which precludes the prosecutor from making a ‘bookmakers’ assessment based on similar cases, sits inconsistently with the case of *R v Clark* – as discussed in Chapter 7 – which allowed evidence of other parental infanticide rates and ‘sudden infant death syndrome’ (SIDS) rates to be compared with each other and used as evidence in the case at trial.

The process which information becomes ‘evidence’ has a number of stages. Firstly, the information is witnessed by someone at the scene. This might be: (a) an eyewitness; or (b) an expert instructed by the police to collect crime scene samples left by the attacker. Then, this information is assessed by the witness. The information is passed to the police, and then it is passed to the CPS. After evaluation by the CPS, the information is passed to the jury for assessment as to its weight and accuracy, and finally a verdict is passed if the jury feel that

there is enough evidence to convict. The passing of this information from witness (either eyewitness or expert), to police, to CPS, to jury, inevitably carries a degree of uncertainty of its accuracy and/or subjective interpretation as to its weight.

The CPS decision to prosecute must ask whether a hypothetical jury would ‘more likely than not’ convict on the strength of the entire case – prosecution *and* defence – facing the defendant. This does not mean that there *is* enough evidence to convict, only that the hypothetical jury’s *perception* of the weight of the evidence is such that there is a likelihood of conviction. Therefore, the test actually asks not whether there is enough evidence to convict, but whether a hypothetical jury *will* convict, based on sound received judicial directions. There are therefore *two* normative elements to the test; (i) that the trial judge will be capable of identifying, and does identify, faulty reasoning in the presented evidence; and (iii) that the jury understands the judicial direction on the presented evidence, can apply it, and does, objectively, impartially and reasonably. The problem, is that the theory and the practice has not been shown to correlate and puts the defendant in a very precarious position if he must rely on sound probabilistic reasoning from the witness(es), the judge and the lay jury – ie a type of reasoning which is routinely found to be faulty.

For instance, it is well known that a committal of the prosecutor’s fallacy [15] has the capability of greatly exaggerating weight of the prosecution case by transposing $P(E|H_d)$ with $P(H_d|E)$. In the case of DNA evidence with a random match probability (RMP) of 1/1 billion and a prior of 1/WP (a world population assumed to be around 7 billion people) this would have the effect of exaggerating the prosecution case as follows:

$P(E|H_d) = 1/1 \text{ billion}$, which is confused with $P(H_d|E) = 1/1 \text{ billion}$

...when actually, $P(H_d|E) \approx 1/7$, meaning that the evidence actually supports the *defence* hypothesis by a factor of 7:1, rather than the *prosecution* hypothesis by a factor of 1 billion:1.

In this example, a juror may convict due to not identifying the flaw in the reasoning, although the evidence would pass the evidential stage test. This routine mis-valuing of probability calculations has been well documented [72] in seasoned medical practitioners who *should* be used to dealing with such matters; so the likelihood of jurors or judges in the legal field - who are unused to interpreting evidence presented in mathematical form - making similar

mistakes, is very high. What is worse is that a Bayesian approach to evidence has been barred from general use by the Court of Appeal in the case of *R v T* [2] - which has meant that an RMP would be routinely presented without instructions as to its probativity in the case as a whole. Where evidence is presented as a RMP, the LR is:

$$P(E|H_p) = 1 \text{ (assuming a 0 error probability rate)} / P(E|H_d) = \text{the RMP}$$

This could give a very high LR in favour of H_p , if the RMP is very small. The problem is that even if the LR greatly favours H_p , the evidence could *still* favour the defence case overall, after it is treated to the full Bayesian model. The CPS advocates that DNA evidence can and should be combined with ‘other types’ of evidence [7].

Since only RMP can be combined with RMP evidence (see Chapter 6 for detailed discussion of this point), it is easy to see how even if DNA evidence is combined with other types of trace evidence, that the case can still favour the defence, even if the weight of accumulated evidence *seems* to favour the prosecution. In these instances a Bayesian model would act as an aid to a transparent and efficient decision-making process.

There are three opportunities throughout the trial process to use a Bayesian model and compare results:

- (i) By the lay or expert witness himself, pre-trial. The witness will place values based on his perception of the weight of the evidence. For instance, if the evidence is an eyewitness account of a physical trait of the suspect, this can be converted to form a RMP, much in the same way a DNA sample can be valued.
- (ii) By the CPS, pre-trial. After receiving the evidence from the witness, the prosecutor can adjust the value of the evidence, based on what he feels the jury’s perception of the weight of the evidence would be.
- (iii) By the jury, post-trial. After hearing the evidence in the case, the jury will have the opportunity to re-value the witness’s evidence after hearing the whole case. This preserves the autonomous role of the trial fact finder, of ascribing weight to evidence [51].

The use of model (ii) will show that the CPS has made a rational decision to prosecute, based on the foundation framework provided by model (i). Model (iii) will expose any biases and fallacies the jury might have in interpreting the evidence. For example, if the evidence in Model (i) is DNA with a RMP of 1/1 billion, this (in the absence of *all* other evidence) would support the defence hypothesis. Once the prosecutor has adjusted the weight of this evidence in model (ii), the jury will have the opportunity to see these evaluations and then apply his own valuation. The whole system will now run as transparently and efficiently as possible, not allowing for biases and fallacies to pervade the process, while also allowing a defendant the opportunity to know exactly the cases facing him in case he wishes to challenge the decision to prosecute.

Of course, an alternative to Models (i) and (ii) would be to ensure that judges and juries are properly trained in the art of proficient routine probabilistic reasoning, perhaps using BNs during or after the trial. However, since this is unlikely, due to (a) the reluctance of the legal community to embrace commonplace Bayesian reasoning; and (b) the lack of a guarantee that mid or post-trial BNs would be correctly constructed or applied, the pre-trial Models act as a safeguard against later problems. In any case, there is no reason not to use the three Models as a ‘compare and contract’ mechanism between the actors in the case: the witness(es), CPS and jury.

Currently, the decision to prosecute is not open to effective redress, since the prosecutor’s own subjective valuation of the evidence is not considered. All the prosecutor need do is argue that he has *considered* the fact finder’s likely re-evaluation of the evidence, and his decision to prosecute is vindicated, which, without a formal framework to show this decision-making process, will mean that a successful case against a prosecutor is unlikely to be forthcoming. Therefore, Codes 3.4 and 3.5 have little to no effect in cases where there is little prosecution evidence against the accused, but the prosecutor believes that the fact finder will more likely than not convict anyway.

8.4 Submission of ‘no case to answer’:

The submission of no case to answer (SNCA) allows a judge to stop a trial at any point for lack of evidence, on the basis that unmeritorious cases will be saved from unnecessary public

expense and defendants from going through an unnecessary trial. The test comes from the case of *R v Galbraith* [73]: ‘...where the judge concludes that the prosecution evidence, taken at its highest is such that a jury properly directed could not properly convict on it, it is his duty, on a submission being made, to stop the case.’ What ‘taken at its highest’ means, is ‘if the prosecution evidence is to be believed’ – ie if it has not been refuted by the defence. This is a different test to the decision to prosecute, which allows for both the prosecution *and* defence cases to be considered (but by the prosecutor pre-trial, not the judge during trial), and is therefore a lower test.

The problem with this test is that regardless of the test’s objective nature, it is largely at the judge’s discretion as to how he values the prosecution case, and without a formal framework with which to gauge the accuracy of this valuation is very difficult for a defendant, believing his has been wronged, to found a successful claim against the judge for applying the incorrect evidential weight. In fact, as with the prosecutor’s pre-trial decision to prosecute, the only thing that the judge need do is show that he has considered the prosecution evidence on a submission of no case to answer, and his decision to proceed with the trial will be vindicated.

The appeal courts, in deciding whether the judge has made the correct decision, will unlikely interfere with the decision, unless the test has been incorrectly applied, or the decision has not been made at all. The reason is that the appeal will be on a point of law or procedure – while the appellate court is free to ask itself whether there was sufficient evidence to allow the case to proceed, it will not be a re-trial where the facts of the case will be open to a new interpretation by the upper courts. In any case, the appellate judge’s decision, in using the same test, will be as potentially flawed as the first instance trial judge’s, such as was applied in Barry George’s compensation claim [74] against the Secretary of State for justice for wrongly bringing his re-trial. As the High Court observed: ‘*[The counsel for the defendant] did submit that the evidence left to the jury was weak and that although four planks of evidence were advanced by the Crown as we have identified above, in fact there was only one: the identification by the witness Mayes.*’ The question which the court must ask itself is: “*Has the claimant established, beyond reasonable doubt, that no reasonable jury (or magistrates) properly directed as to the law, could convict on the evidence now to be considered?*” In light of the discussion in Chapter 5, which explains the subjective nature of the criminal standard of proof (BARD), this test contains no formalised framework with which to provide adequate redress for the wrongly-tried defendant.

In George's compensation claim following his re-trial acquittal in *R v George* [8], the High Court rejected the submission that there was no evidence to proceed with the case after the firearm discharge residue (FDR) evidence, which had been presented as part of the prosecution case in the defendant's first trial, had been excluded from his re-trial: '*... in the end, there was no submission formulated by [counsel for Barry George] capable of persuading us that the trial judge was wrong to leave the case to the jury. In our view, this has the consequence that this claimant's case inevitably fails the test even as formulated by [counsel for Barry George]*'. While this judgment is not conclusive that any claim would have failed, the words '*there was no submission formulated, capable of persuading us that the trial judge was wrong*' is indicative of the fact that the appeal courts are (a) reticent to delve deeply into the trial judge's decision-making process; and (b) have no means of providing actual proof that the decision made by the trial judge, and later itself, is sound.

In September 2013, Keir Starmer, Director of Public Prosecutions (DPP), in addressing a Home Affairs Select Committee [75], defended the CPS's pre-trial decision to prosecute the actor Michael Le Vell for rape, and stated that the correct procedure for the decision to prosecute had been followed and test applied, and during the trial, the judge had correctly applied the test for a submission of no case to answer. This indicates that the DPP, as well as the appellate courts, are reluctant to examine the substance of the trial judge's decision, as long as the correct procedure in applying the test has been followed.

8.5 Remedies available to the wronged defendant:

The remedy for a defendant who wishes to bring an action against the CPS for a wrongful decision to prosecute, or against a judge who has wrongly allowed a trial to continue, comes by way of judicial review. The action is not an appeal, but a challenge to the manner in which a decision by a public authority has been made.

Part 54 of the Civil Procedure Rules 1998 [76] provides the definition for judicial review:

54.1(2)(a) a "*claim for judicial review*" means a claim to review the lawfulness of-

(i) an enactment; or

(ii) *a decision, action or failure to act in relation to the exercise of a public function.*

However, there is little hope of challenging the prosecutor's decision to prosecute by way of judicial review, since as Blackstone's Criminal Practice 2011 [77] suggests: '*...it may be challenged within the trial process itself, notably by an application to stay proceedings on the grounds of abuse of process.*' The basic principle is that it is for the prosecution, not the court, to decide whether a prosecution should be commenced and, if commenced, whether it should continue. However, the courts have an overriding duty to promote justice and prevent injustice. [78]. 'Abuse of process' has been defined as something so unfair and wrong with the prosecution that the court should not allow a prosecutor to proceed with what is, in all other respects, a perfectly supportable case [79]. 'Unfair and wrong' is for the court to determine on the individual facts of each case.

The categories for a stay for an abuse of process are not limited, but the authorities which may apply in cases relating to lack of evidence, are largely (i) An abuse of executive power; and (ii) Failing to obtain evidence. There are problems, however, with both of these categories if applied to the decision to prosecute:

8.5.1 An abuse of executive power:

This category includes cases such as *Bennett* [80] where the police have abducted the suspect and brought him to trial without complying with the correct arrest and detention procedure. In *Redmond* [81] it was decided that the crucial issue was whether the prosecuting authorities have knowingly abused their powers.

In cases where there has been a wrong decision to prosecute, the malice of the prosecutor is rarely an issue. In *George*, it was simply that the evidence was of neutral value – the prosecutor's state of mind may have been as neutral as the evidence.

8.5.2 Failing to obtain evidence:

This category normally relates to material which was present at the decision to prosecute stage, but has somehow gone missing before trial. In *Dobson* [82], the test was stated to be

whether the police had a duty to obtain evidence, and whether they have, in bad faith, failed to obtain it. The Evidential Stage test at paragraph 4.5 contains the duty to obtain evidence, but even though there may have been a failure to obtain enough evidence to prosecute, the ‘bad faith’ element is more often than not absent to allow a stay for an abuse of process.

In *Ebrahim* [83], it was decided that there must be a serious fault on the part of the prosecution authorities for an application for a stay to succeed. It is so far untested as to whether not collating enough evidence to pass the Evidential Stage test is a ‘serious fault.’

In any case, a claim for judicial review is unlikely to succeed pre-trial, thereby ensuring that a wronged defendant actually stands trial to attempt to argue that he should not be there in the first place. The inherent jurisdiction of the court to stop a prosecution to prevent an abuse of process is to be exercised only in exceptional circumstances [78]. In *R(B) v DPP* [68], Toulson LJ confirmed the position: *‘The exercise of the court’s power of judicial review is less rare in the case of a decision not to prosecute than a decision to prosecute (because a decision not to prosecute is final, subject to judicial review, whereas a decision to prosecute leaves the defendant free to challenge the prosecution’s case in the usual way through the criminal court).’* This approach is expensive and time consuming, and can be incredibly damaging for a defendant who should be entirely free of trial procedure if the prosecution case against him is weak. The problem, is that the essential focus of the doctrine is on preventing unfairness at trial (not pre-trial), through which the defendant is prejudiced in the presentation of (not whether he should be in the position of defending), his or her case.

The implementation of a Bayesian model pre- and/or mid-trial would reduce the emotional toll on defendants, who, by rights, should not be tried; since as Annabelle Jones [24] observes: *‘If Barry George is innocent of the murder of Jill Dando and was in fact convicted on the [62] basis of the jury’s misunderstanding of the forensic evidence, he has spent nearly eight years of his life behind bars for a crime that he did not commit. The effects of this on any individual would be long-lasting; the effects on an arguably already unstable and vulnerable individual such as Barry George could be even more damaging.’* Of course, Jones’ comment that the defendant ‘did not commit the crime’ was simply an allusion to his acquittal and not whether the FDR evidence in the case actually DID have probative effect towards the prosecution hypothesis – for further discussion on this point Chapter 7 discusses the FDR evidence, and the two papers written by Fenton, Berger et al [10] [62] provides a full Bayesian model to explain and expose the flaws of the reasoning in the case.

8.6 Conclusion:

Having explored in detail the Full Code Test and SNCA test thresholds, it is obvious that the large amount of discretion afforded the decision-makers – the CPS and the trial judges in the case – in the decisions of whether to bring or continue the case, pre- and/or mid-trial, means that the wrongly-tried defendant has very little chance of success in appealing these decisions if there is no formalised way to show how these decisions were arrived upon.

Take a simple example: The police have charged the Defendant with one count of stabbing with intent to kill. Bayes theorem shows that there is only a 50% chance of the Defendant's guilt – perhaps on DNA match evidence with a fairly high RMP, combined with eyewitness evidence of a reasonably common physical trait - which, absurdly, would be enough to allow the case to proceed to trial if a jury mistakenly believes that the evidence is enough to convict.

Let us say that, on presentation of the case, the jury labouring under biases and/or fallacious reasoning, believes that there is a more than 90% chance that the Defendant is the attacker and that this is a high enough threshold to put the prosecution hypothesis beyond reasonable doubt. The CPS, in full knowledge that there is only a 50% likelihood of guilt, would be forced to proceed with a trial.

Therefore, it is proposed that a scale of guilt would be presented along with the evidence in the case, for the jury to choose:

> 50%

50 – 60%

60 – 70%

70 – 80%

----- Proposed threshold

80 – 90%

90 – 99%

99 – 100%

If more than 50% of the jury choose, say, at least 80 – 90%, then the CPS would have to proceed on the basis that it is more likely than not that the jury would convict.

But surely in the nature of transparency and the overriding justice objective in criminal trials, the CPS would be compelled to disclose the counter-intuitive errors of the jury's decision? With that being the case, it would surely be better for the BT calculations to be disclosed as part of the pre-trial evidence, rather than allow the jury to labour under a misapprehension as to the value of the presented evidence.

As far as criticism that 80 – 90% does not represent a threshold which would indicate a juror would convict if passed, is concerned, the argument would have to be posed: what figure would suffice? Obviously, a juror intuitively decides whether a threshold, by which to convict, has been passed, otherwise only a 100% certainty standard of proof would do. So therefore, an arbitrary line must be drawn, and 80% has been chosen.

Conversely, if when all of the evidence has been calculated and there is at least a likelihood of around 80 – 90% in support of the guilty hypotheses, but a lay person only believes that the evidence has a small chance of convicting the Defendant, does the CPS proceed? The answer would have to be 'no'.

Therefore, within this chapter, we recommend a Bayesian approach to pre- and/or mid-trial decision-making which will have various positive effects including:

- Providing a transparent, logical, formalised framework approach to pre- and/or mid-trial decisions in order to save time, expense and the risk of a wrongful conviction of defendants by juries with natural biases against tried defendants - even if the weight of evidence does not merit a conviction.
- Allowing an appeal court the proper tools necessary to decide whether or not to allow an appeal for a wrongly-brought trial, thereby saving time and expense.

- Allowing a defendant the tools necessary to decide whether or not his claim for a wrongly brought case has merit in an appeal, thereby saving time and expense.

All of these points support not only the research hypothesis, but also help generate debate and make necessary and vital recommendations in reforming the Full Code Test and SNCA process.

9 CHAPTER IX: Overall Conclusions

The research hypothesis is: *'It is inevitable that in criminal cases with multiple pieces of evidence, a Bayesian approach to evidence evaluation and presentation must be used'*. The main areas of research, therefore, were:

- An investigation of the need for probabilistic modelling, whether pre- or mid-trial
- The amenability and feasibility of evidence to modelling
- A question of whether Bayes' theorem (BT) is the most appropriate method to model uncertainty
- Whether the current methods of explicit probabilistic modelling are correct and appropriate

The approach taken in this thesis was to research the problems with probabilistic reasoning in the legal cases such as those listed in Chapter 4 and examine the reasons why these problems seemed to routinely occur. The conclusions drawn were complex and varied. After an initial investigation, with reference to industry-recognised practitioner materials by Buckleton and the Nuffield Foundation, it was obvious that there were many inconsistencies in the approach to the foundation principles of probability theory which had the potential to lead to the wrong weight being attributed to pieces of evidence in the case, thereby leading to the natural conclusion of the entire prosecution hypothesis - in cases such as *R v Sally Clark* - or defence hypothesis - in cases such as *R v Barry George* - being unfairly weighted.

This disparity was evident in the way that fundamental probabilistic principles, such as those discussed in Chapter 2, are applied by practitioners and expert witnesses to complex legal cases. One of the most common fallacies committed is (and will continue to be, without BT applied to negate its effects) is the 'prosecutor's fallacy'. However, this fallacy is only considered common because it is fairly easy to identify due to its explicit probabilistic presentation and its dramatic effect to a prosecution or defence case.

Chapter 3 began by examining the prosecutor's fallacy and the problems for legal hypotheses when one occurs, and examined the UK courts' attitude towards the fallacies. This line of research uncovered a worrying misinterpretation of the fallacy by the courts, which only after

a Bayesian analysis of the problem exposed the issue that there could never be an ‘acceptable’ committal of the fallacy, regardless of the Court of Appeal ruling on the matter. An application of BT to the problem would have eradicated this and would prevent this problem from occurring in the future. This chapter looked at the *need* for BT and concluded that there is a need.

While examining the possibility of using BT to eradicate the prosecutor’s fallacy, the research area began to expand to consider the use of BT to examine other potential fallacies and misinterpretations of the probative value of probabilistic evidence. The first task was to examine why some types of evidence are considered by the UK courts in rulings such as *R v T* (2010) to be ‘probabilistic’ (and therefore amenable to BT) and why others are not. This distinction is researched in Chapter 5 and the conclusions drawn are simple – that there is no distinction between different types of trace evidence; which means that, theoretically, all trace evidence is capable of being modelled in a Bayesian network (BN) for evaluation pre-trial or in-trial presentation to a jury.

Chapter 5 also examined the key alternative method to BT, in order that any recommendation of BT stands up to scrutiny as the best and most appropriate method of evaluating and presenting evidence. The Wigmorean method – a method recommended by legal-mathematical scholars, supported by the Nuffield Foundation in London – is discussed and examined. This chapter concludes the following: (a) that as far as the *amenability* of evidence to modelling is concerned, that the evidence is amenable; and (b) that as far as the *appropriateness* of BT for use as the recommended method of evaluating and presenting explicitly quantified evidence in a case is concerned, that BT is the most appropriate method.

Chapter 6 contains the most in-depth study of prior probabilities ever conducted. Until now, in most academic studies, the only controversy exposed by a discussion of priors is that of their subjective, and therefore uncertain, nature – an issue which UK courts finds unpalatable. However, what the research in this chapter discovered was that the value ‘*K*’ – denoting context - in mathematical models, is routinely ignored. This crucial mistake can, and does, lead to miscalculations of the probative value of evidence, due to consideration of the existence of causal pathways (as discussed in Chapter 2) which do not exist in the way the experts believed they did. These problems can only be rectified by a Bayesian approach to the cases as a whole. Interestingly, Chapter 6 does not deal with explicit ‘fallacies’ in the way

that Chapter 3 does, but this problem with K potentially far more damaging, as the problems are buried within the model, would likely never be disputed by prosecution or defence counsel or judges, and would never come to light without the discussion in this chapter. In fact the Crown prosecution Service (CPS) guidelines, which govern the way cases are brought and presented, overlook this problem at the most fundamental level. Therefore, while in this thesis BT is recommended to eradicate fallacies, it also becomes the vital tool in eradicating basic errors in probabilistic reasoning connected with all aspects of evidence in a case.

In fact, the conclusion drawn in this chapter is a complicated one, as can be seen in paragraph 6.13. In essence, it is that BT should be used to eradicate the fallacy that evidence across types can be combined using the traditional Bayesian methods. This is done by examining the hypothesis pairing which relates to the individual piece of presented evidence, and ensuring that K is not compromised when that piece of evidence is combined with the other evidence in the case.

Chapter 7 examined the likelihood ratio (LR) and discusses how the full Bayesian approach to probabilistic reasoning must be used, instead of the ‘lone LR’ approach favoured by some academics. This chapter builds on the recent papers published by Fenton, Berger et al which centre on problems within the *R v George* and *R v Clark* cases, and concludes with the recommendation that in order for a piece of evidence to carry the correct probative value, it must be considered in light of all of the evidence in the case – the method for which must be a full Bayesian model.

Chapter 8 referred back to the CPS guidelines - which were first analysed in Chapter 6 - and explains how BT must be used, pre-trial, to assist prosecutors in whether or not to proceed with a case to trial by calculating, probabilistically, the probative value of the evidence in the case towards the prosecution hypothesis. The chapter concludes that without this formalised approach to evidence evaluation, that there is no appropriate means of redress for wrongly-tried defendants.

On a more general level, the author asserts that BT is the only means of quantifying uncertainty, exposing unreasonable assumptions and charting dependencies between supposed separate pieces of evidence in complex legal cases, in order that any decision-

making – whether by prosecutors, defence counsel, juries or judges - pertaining to a case, is accurate, transparent and logical. BT is not merely a method of quantified mathematical modelling, it helps us (a) understand the relationship between pieces of evidence and their causal pathways to a given hypothesis; (b) recognise the crucial (and oft-forgotten) issue of context; and (c) identify and negate our natural biases and intuitions which may otherwise have – and undoubtedly routinely has - clouded our judgement.

Acknowledgments:

I would like to thank the following for their invaluable contributions, in various ways, in helping me to complete this thesis: Norman Fenton, Martin Neil, Richard Nobles, Ian Dennis, Carole McCartney, Charles Wild, Anne Hsu, Eugene Dementiev, David Lagnado and the Engineering and Physical Sciences Research Council (EPSRC).

Bibliography

- [1] "R v Doheny & Adams," [1997] 1 Cr App R 369.
- [2] "R v T," 2010 EWCA Crim 2439.
- [3] N. Fenton and M. Neil, Risk assessment and decision analysis with Bayesian networks, London: CRC Press, 2013.
- [4] L. Tribe, "Trial by Mathematics: Precision and ritual in the legal process," *Harv L Rev* 1329,, vol. 84, pp. 1361-5, 1971.
- [5] "People v Collins," 1968 68 Cal. 2d 319 .
- [6] J. S. Buckleton, S. J. Walsh and C. Triggs, Forensic DNA Evidence Interpretation, CRC Press (an imprint of Taylor & Francis), 2005.
- [7] Forensic Science Service Guide to DNA for lawyers and investigating officers, FSS Communications Dept, 2004.
- [8] "R v Barry George," [2007] EWCA Crim 2722.
- [9] "Regina v Sally Clark," [2000] WL 1421196.
- [10] N. Fenton, D. Berger et al, When 'neutral' evidence still has probative value: implications from the Barry George Case, vol. 54, Science & Justice, 2014, pp. 274-287.
- [11] N. Fenton, M. Neil and A. Hsu, "Presentation of Bayes in Court," 2012 as yet unpublished.
- [12] W. Edwards, "Influence Diagrams, Bayesian Imperialism, and the Collins Case: An appeal to Reason," vol. 13, p. 1025, 1991.
- [13] "C v The Queen," [2011] EWCA Crim 1607.
- [14] "R v Kelly Gray," [2005] EWCA Crim 3564.
- [15] W. Thompson and E. Schumann, "Interpretation of statistical evidence in criminal trials - the prosecutor's fallacy and the defence attorney's fallacy.," *Law Hum Behav*, vol. 11, pp. 167-187, 1987.
- [16] P. Roberts, C. Aitken and G. Jackson, "Fundamentals of Probability and Statistical Evidence in Criminal Proceedings (Nuffield Practitioner Guide No.1)," *Royal Statistical Society's Working Group on Statistics and Law*, 2010.
- [17] J. Curran, "An introduction to Bayesian credible intervals for sampling," *Law, Probability and*

- Risk*, vol. 4, p. 115–126, (2005).
- [18] M. Redmayne, "Doubts and burdens: DNA evidence, probability and the courts," *Criminal Law Review*, p. 464, 1995.
 - [19] R. Puch-Solis, P. Roberts, S. Pope and C. Aitken, "Assessing the probative value of DNA evidence (Nuffield Practitioner guide No.2)," *Royal Statistical Society's Working Group on Statistics and the Law*, 2011.
 - [20] C. Colmez and L. Schneps, *Maths on Trial*, Basic Books, 2013.
 - [21] J. Bernoulli, "Ars Conjectandi: Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis," 1713.
 - [22] "People v Puckett," [2008] No. A121368, Cal. Ct. App., 1st Dist., May 1, .
 - [23] M. Finkelstein and W. Fairley, "A Bayesian Approach to Identification Evidence," *Harv LR*, vol. 83, p. 489, 1971.
 - [24] A. Jones, "Fresh evidence: probative value of firearms discharge.," *J. Crim. L.* , vol. 72, no. 3, pp. 199-201, 2008, .
 - [25] Law Commission Report: Expert Evidence in Criminal Proceedings in England and Wales, London Stationery Office, 325/2011.
 - [26] R. Friedman, "Answering the Bayesioskeptical challenge," *E & P*, vol. 1 (Spe), pp. 276-291, 1997.
 - [27] C. Aiken et al, "Expressing evaluative opinions: A position statement," *Science & Justice*, vol. 51, no. 1, pp. 1-2, 2011.
 - [28] R. Allen, "Rationality, algorithms and juridical proof: A preliminary enquiry," vol. 1(Spe), pp. 254-275, 1997.
 - [29] J. Bring, "Bayes on Trial," *E & P*, vol. 1(Spe), pp. 292-295, 1997.
 - [30] P. Roberts, C. Aitken and G. Jackson, "Case Assessment and Interpretation of Expert Evidence (Nuffield Practitioner Guide No.4)," *Royal Statistical Society's Working Group on Statistics and the Law*, 2013.
 - [31] P. Tillers and J. Gottfried, "Case Comment - United States v Copeland: A collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable?," *Law Probability and Risk*, vol. 5, pp. 135-157, 2006.
 - [32] B. Robinson and A. Vignaux, "Explaining Evidence Logically," *NLJ* 148, pp. 159-162, 1998.
 - [33] A. Biedermann, P. Garbolino and F. Taroni, "'The subjectivist interpretation of probability and the problem of individualisation in forensic science'," *Science & Justice*, vol. 53(2), pp. 192-200,

2013.

- [34] E. S. Neumann, "Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm," *JR Statist Soc A*, vol. 175, no. Part 2, pp. p1-26, 2012.
- [35] P. Roberts, C. Aitken, F. Taroni and A. Biederman, "The Logic of Forensic Proof: Inferential Reasoning in Criminal Evidence and Forensic Science (Nuffield Practitioner Guide No.3)," *Royal Statistics Society's Working Group on Statistics and the Law*, 2013.
- [36] J. Kadane and M. Redmayne, "Probabilistic evaluation of the Sacco and Vanzetti evidence," *International Journal of Evidence & proof*, vol. 1(2), p. 177, 1997.
- [37] Crown Court Benchbook 'Directing the Jury', Judicial Studies Board, 2010.
- [38] C. Aitken and F. Taroni, "Fundamentals of statistical evidence: a primer for legal professionals," *International Journal of Evidence & Proof*, vol. 12(3), p. 181, 2008.
- [39] G. Jackson, Champod and I. Evett, "Principles of interpretation-application of the likelihood ratio in marks cases," *FSS Paper to European Shoe Print & Tool Mark Examiners*, 2001.
- [40] M. Redmayne, P. Roberts, C. Aitken and G. Jackson, "Forensic science evidence in questions," *Crim LR*, vol. 5, pp. 347-356, 2011.
- [41] L. Laudan, *Truth, Error and the Criminal Law – an essay in legal epistemology*, Cambridge University Press, 2006.
- [42] J. Franklin, "Case comment – United States v Copeland: quantification of the 'proof beyond reasonable doubt' standard," *Law Probability and Risk*, vol. 5, pp. 159-165, 2006.
- [43] Archbold Criminal Pleading Evidence and Practice, Sweet & Maxwell, 2015.
- [44] "R v Summers," (1952) 36 Cr. App. R. 14.
- [45] "R v Bracewell," (1979) 68 Cr. App. R. 44.
- [46] "R v Adams (Denis)," [1996] 2 Cr App R 467.
- [47] N. Fenton and M. Neil, "On limiting the use of Bayes in presenting forensic evidence: an irrational but understandable ruling," Extended draft, Available at: https://www.eecs.qmul.ac.uk/~norman/papers/likelihood_ratio.pdf, 2011.
- [48] R. Allen, "Rationality, algorithms and juridical proof: A preliminary enquiry," *E & P*, vol. 1(Spe), pp. 254-275, 1997.
- [49] C. Jowett, "Sittin' in the Dock with the Bayes," *NLJ*, vol. 151, p. 201, 2001.
- [50] P. Tillers, "Trial by mathematics - reconsidered," *Law Probability & Risk*, vol. 10, pp. 167-173,

2011.

- [51] Schum and Martin, *Handbook of Applied Cognition*, John Wiley and Sons Ltd, 2007.
- [52] J. Bring, "Bayes on trial," vol. 1(Spe), pp. 292-295, 1997.
- [53] J. Berger, "The case for objective Bayesian analysis," vol. 1, no. 3, pp. 385-402, 2006.
- [54] "R v Adams (Denis)," [1996] 2 Cr App R 467.
- [55] I. Dror and G. Hampikian, "Subjectivity and bias in forensic DNA mixture interpretation," *Science & Justice*, vol. 51, pp. 2014-208, 2011.
- [56] D. Balding, "Weight of Evidence for Low-Copy Number DNA Profiles," *QMUL seminar*, 29/2/12.
- [57] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124-1130, 1974.
- [58] "R v Whiting," [2010] EWHC 1329 (QB).
- [59] M. Redmayne, *Expert Evidence and Criminal Justice*, London: Oxford University Press, 2001.
- [60] N. Fenton, Model date: 30/04/14.
- [61] C. Vlek et al, "Building Bayesian Networks for Legal Evidence with Narratives," *Artificial Intelligence & Law*, vol. 22(4), pp. 375-421, 2014.
- [62] N. Fenton, D. Berger et al, "Response to "On the use of the likelihood ratio for forensic evaluation: response to Fenton et al"," *Science & Justice*, vol. 54, pp. 319-320, 2014.
- [63] I. Evett, R. Cook, G. Jackson, P. Jones and J. Lambert, "A hierarchy of propositions: deciding which level to address in casework," *Science and Justice*, vol. 38, pp. 231-239, 1998.
- [64] P. Dawid, "Probability and Proof. Online Appendix to "Analysis of Evidence" T. Anderson, D. A. Schum and W. Twining.," Cambridge University Press, 2005.
- [65] N. Fenton, "Assessing Evidence and Testing Appropriate Hypotheses," *Science & Justice*, vol. 54(6), pp. 502-504, 2014.
- [66] "Nulty v Milton Keynes," [2013] EWCA Civ 15.
- [67] "The Popi M," [1985] 1 WLR 948.
- [68] "R (B) v DPP," [2009] EWHC 106 (Admin).
- [69] DPP, *Full Code Test for Crown Prosecutors*, Crown Office, 2010.

- [70] "Bayes and Law Consortium," London, 2012.
- [71] D. Tversky and A. Kanheman, "Prospect Theory: An analysis of decisions under risk," *Econometrica*, vol. 47, no. 2, pp. 263-291, 1979.
- [72] W. Casscells et al, "Interpretation by physicians of clinical laboratory results," *NEJM*, vol. 299, no. 18, p. 999, 1978.
- [73] "R v Galbraith," [1981] 1 WLR 1039.
- [74] "R (George & others) v Secretary of State for Justice," [2013] EWHC 72 (Admin).
- [75] "Michael Le Vell 'celebrity witch-hunt' suggestion denied," BBC News, Manchester, 2013.
- [76] Civil Procedure Rules, vol. 1, Sweet & Maxwell , 2011.
- [77] Blackstone's Criminal Practice, Oxford University Press, [2011].
- [78] "Crown Prosecution Service, Prosecution Policy and Guidance, Legal Guidance" 2015.
- [79] "Hui Chi-Ming v R [1992] 1 A.C. 34, PC".
- [80] "Horseferry Road Magistrates' Court, ex parte Bennett [1994] 1 AC 42".
- [81] "R v Redmond [2001] EWCA Crim 1606".
- [82] "R v Dobson," [2001] EWCA Crim 1606.
- [83] "R (Ebrahim) v Feltham Magistrate's Court [2001] 1 WLR 1293".